sociological science

Supplement to:

González-Bailón, Sandra, David Lazer, Pablo Barberá et al. 2024. "The Diffusion and Reach of (Mis)Information on Facebook During the U.S. 2020 Election" Sociological Science 11: 1124-1146.

Supplemental Materials

The Diffusion and Reach of (Mis)Information on Facebook during the US 2020 Election

Sandra González-Bailón, David Lazer, Pablo Barberá, William Godel, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M. Guess, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, Joshua A. Tucker

Correspondence to: sandra.gonzalez.bailon@asc.upenn.edu

Contents

S1 Overall Project Description

S1.1 Meta-Academic Collaboration Overview

This paper is part of the *US 2020 Facebook and Instagram Election Study*, a broader set of experimental and observational studies that occurred as a result of a collaboration between academics and Meta. The project was designed to address three intertwined concerns related to scientific understanding of the impact of social media on democratic processes. First, in the aftermath of the 2016 US elections, there was a widely recognized need to understand the impact of social media platforms on US elections. Second, research conducted solely by platform employees could encounter skepticism from the mass public and policy community. And third, outside independent researchers not employed by the platforms faced legal and fiduciary challenges in securing access to the data and research pipelines necessary to conduct rigorous scientific analyses about the impact of social media on elections.

The *US 2020 Facebook and Instagram Election Study* is an attempted solution to this bundle of challenges. The project represents a novel form of collaboration between a team of researchers at Meta and a set of external researchers.[S1] The costs associated with the research (e.g., participant fees, recruitment, data collection, etc. – see section S2.2) were paid by Meta. The academic team members received no form of financial or any other compensation (e.g., support for student assistants, course buyouts, research funds) from Meta for their participation in the project (see S1.2 for other measures applied to ensure the research integrity of the project).

Genealogy of the Project In early 2020, researchers at Facebook (now Meta) approached Social Science One[S2] about the possibility of jointly organizing a research project around studying the impact of Facebook and Instagram on the November 2020 US elections. Social Science One had been created to facilitate industry-academia collaboration to study social media platforms and their impact on society and, in particular, to make data available to researchers who did not work for Meta. Social Science One at that point consisted of two directors/founders (Gary King and Nate Persily) and a series of advisory committees, each of which had a chair. These included Professor Natalie Jomini Stroud, the Chair of the North America Advisory Committee, and Professor Joshua A. Tucker, the Chair of the Disinformation and Electoral Integrity Committee. As the Chairs of the Social Science One Committees that were most related to the proposed project, Stroud and Tucker agreed to jointly co-Chair the academic team[S3] to collaborate with Meta on what has come to be called the *US 2020 Facebook and Instagram Election Study*.

In the interest of balancing the competing needs of assembling a research team quickly and ensuring the necessary research expertise for the project, Stroud and Tucker made the decision

[S1] At the time the project began in the spring of 2020, the company involved was called Facebook. For the sake of simplicity, we refer to the company by its current name, Meta, in the rest of the supplemental materials.
[S2] https://socialscience.one.
[S3] https://socialscience.one/blog/new-data-new-datasets-new-research-projects.

to recruit the remainder of the academic team from existing Social Science One advisory committee members based on their diverse research expertise. In March of 2020, Stroud and Tucker approached 16 of these advisory committee members, 15 of whom agreed to join the academic team for the project. Subsequently, one of the 15 withdrew in the first months of the project because of competing time demands. In addition, one of the advisory committee members requested that a current co-author on very closely related projects and with relevant statistical expertise for the project be allowed to join the team as well, resulting in an academic team of 17 members including Stroud and Tucker[S4]. Meta had no say in who was selected by Stroud and Tucker to be part of the academic team.

Chad Kiewiet de Jonge was the Meta research manager who oversaw day-to-day management of the research project at Meta. Annie Franco and Winter Mason co-led the Meta research team, which grew to include 16 researchers, 2 data engineers, 1 data scientist, and 3 interns working on various parts of the overall project.

Project Execution Once assembled, the team of academics met beginning in March of 2020 to first brainstorm research ideas within the project's mandate of studying Facebook and Instagram's impact in the context of the US 2020 elections and then to develop ideas for specific paper proposals. Concurrently, the team of Meta researchers began working with the academic team to provide feedback on research proposals, including the feasibility of possible designs and procedures for collecting the necessary data. As a result of this process, four general areas of inquiry were selected to form the scope of project: (1) dis/mis/information, knowledge, and (mis)perception; (2) political polarization; (3) political participation, both online and offline, and including vote choice and turnout; and (4) attitudes and beliefs about democratic norms and the legitimacy of democratic institutions.

The next step in the project involved identifying specific paper topics within the four general scope conditions. Based on their research interests, a subset of independent academic researchers served as "core authors" of each paper and were given control rights over final versions of the pre-analysis plans (PAPs) and papers.[S5] Both the academic researchers and the Meta researchers worked together to design the pre-analysis plans.[S6] The core authors for this paper and associated PAP are Sandra González-Bailón and David Lazer.

Data collection was carried out by Meta and, for off-platform data, by NORC, an independent survey research organization at the University of Chicago.[S7] Meta recruited most participants and collected on-platform data, while NORC carried out all surveys associated with the project, collected and appended all supplemental data outside of the Facebook/Instagram on-

[S4] https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election.

[S5] By *control rights*, we mean that in the event of disagreements between members of the research team, the core authors would have the final say in resolving these disagreements.

[S6] A total of 16 pre-analysis plans were registered at https://osf.io/ek29s/registrations..

[S7] NORC was selected following a competitive bidding process involving other online survey research firms. Employees of NORC who implemented the data collection process were not members of the independent academic research team. More details about NORC can be found at https://www.norc.org/Pages/default.aspx.

platform data, and recruited additional survey panelists (see section S2 for more details). The academic research team did not contact any human subjects as part of the research efforts. In the rare cases where members of the academic team – whose names had been publicly announced[S8] – were messaged by study participants, the messages were passed to NORC to respond.

At the data analysis stage, the Meta team produced, and the academics reviewed and approved, pipeline code used to produce the data tables needed for this project from raw platform data (e.g. number of followers) and data created for other internal Meta purposes (e.g. predictions of ideology of US Facebook users) that were employed in the analysis. The Meta researchers and, in some instances, the independent academics, carried out the initial analyses as detailed in the pre-analysis plan and as deemed necessary by the full research team for mutually agreed upon research-relevant analyses. The academics' role in the analysis was to contribute to and monitor the results of data analyses conducted by the Meta research team, including: reviewing and, in some cases writing, code; inspecting de-identified samples or aggregated outputs through screen sharing; and, when possible, replicating the analyses within Meta's secure data-sharing Researcher Platform using data that has been stripped of any individually-identifying information.

Drafts of papers were written by the academic research team members, with feedback from the Meta academic researchers but with final control rights resting with the specified core academic authors (in the case of this paper, Sandra González-Bailón and David Lazer).

Figure S1 summarizes the genealogy and execution of the project up to the submission of this paper. A full description of the roles and responsibilities of the academic research team, the Meta researchers, and NORC can be found at the Open Science Foundation repository, currently under embargo.[S9]

S1.2 Research Transparency and Integrity

One of the primary goals in designing the project was to build in transparency concerning the research process given the constraints under which we were operating. With this in mind, five conventions were adopted to guide the research process.

- First, none of the academic researchers nor their institutions received financial or any other compensation (e.g., support for student assistants, course buyouts) from Meta for their participation in the project.
- Second, the analysis for all of the papers resulting from the project, including this one, were pre-registered at the Open Science Foundation (that is, we specified research questions, hypotheses, methods, and planned analysis before data collection). The pre-registrations were embargoed while the research was being carried out, but will be publicly released at time of publication. The pre-analysis plan (PAP) that was registered for this study is

[S8] https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election.
[S9] https://osf.io/7wpgd.

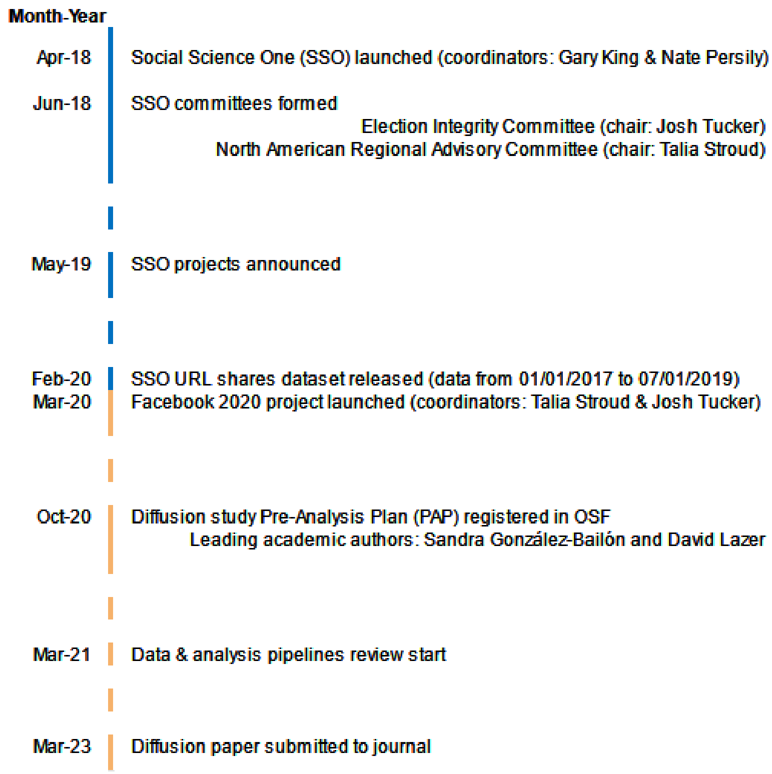

Figure S1: **Project Timeline.** Summary of the genealogy of the *US 2020 Facebook and Instagram Election Study* and key dates in the execution of this paper (pre-registered in October 2020).

included in section S7. Replication materials will be publicly released at time of publication.[S10] A list of deviations from and clarifications of the pre-analysis plan can be found in section S8.

- Third, for every paper, a set of core authors with control rights over the final content of the paper were specified in the pre-analysis plan. These core authors consist only of academic researchers (i.e., not employees of Meta). The core authors with control rights for this paper are Sandra González-Bailón and David Lazer.

- Fourth, Meta publicly[S11] agreed that there would be no pre-publication approval of papers for publication on the basis of their findings. At the time the PAPs were proposed – but before any data analysis was conducted – Meta conducted legal, privacy, and feasibility reviews of the studies. Meta was entitled to review papers prior to publication, but could only request changes to protect confidential or personally identifiable information or to abide by their existing legal obligations.[S12] For this article, Meta requested that we revise our description of race and ethnicity data, and definitions of the "repeat offender" policy. The statements about race and ethnicity are included in pages SM-15 and SM-16. The revised sentences originally stated that Meta does not have any data on the race or ethnicity of its users. Meta requested the clarification that it had limited race/ethnicity data coverage for its US users and it could not be used to inform large scale probability sampling. The statements about the "repeat offender" policy are included in the main text, under the section 'Content Moderation on Facebook'. The revised sentences originally omitted that content fact-checked as "altered" (in addition to that labeled as "false") accumulated "repeat offender" strikes; and that Groups required three strikes (not two) to have the visibility of their subsequent content reduced as of the time of the 2020 election.

- Fifth, we appointed a rapporteur for the project – Professor Michael Wagner of the University of Wisconsin, Madison – who was neither a paid employee of Meta nor a member of the academic research team. The rapporteur was given access to the project researchers, allowed to join project-related meetings, and had access to project documents. The rapporteur is not a co-author on any of the papers resulting from the study, but the expectation is that he will publish both academic and popular press articles assessing the research process itself.

We also decided that our primary approach to releasing results publicly would be upon completion of the peer review process. The peer review process involves having other scholars, who are not authors on the paper, review the work, provide criticism and feedback, and make a recommendation as to whether the scholarship is worthy of publication. The output from this

[S10] Pre-registration for this paper is at https://osf.io/b4xvd/; pre-registrations for the entire project, including supplementary materials to the individual analysis plans, can be found at https://osf.io/ek29s.

[S11] https://research.facebook.com/2020-election-research/

[S12] https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election/

project was never intended to be a "report" on Facebook and Instagram's impact on the 2020 US election, but rather a series of peer-reviewed academic publications addressing scientific questions related to the impact of various aspects of the Facebook and Instagram platforms on the 2020 US election.

Finally, Meta plans to make de-identified datasets from each published study conducted under this initiative and designed in collaboration with the academic team available to the broader research community, so that others can reproduce the analyses and conduct further election studies.

Below we also list declarations from the academic author team. For consistency, we use the following key:

a	Current employee (Meta)
b	Past employee (Meta)
c	Own individual stocks (Meta)
d	Paid consulting work (Meta)
e	Direct research funding from Meta (grant to you as PI or Co-PI)
f	Received an honorarium/fee (from Meta) for attending or hosting an event/serving as outside expert
g	Attended a Meta event where food, travel, or lodging was paid for by the company
h	Current employee (at a related company: Twitter, TikTok, Google/YouTube)
i	Past employee (at a related company)
j	Own individual stocks (at a related company)
k	Paid consulting work (at a related company)
l	Direct research funding from a related company (grant to you as PI or Co-PI)
m	Received an honorarium/fee (from a related company) for attending or hosting an event/serving as outside expert
n	Attended an event (at a related company) where food, travel, or lodging was paid for by the company

Author declarations : Hunt Allcott: Microsoft employee; none of the above. Deen Freelon: g. Matthew Gentzkow: f, g, m, n. Sandra González-Bailón: g, l. Andrew Guess: e, g. Shanto Iyengar: e, g. Young Mie Kim: g. David Lazer: g, n. Neil Malhotra: g, n. Brendan Nyhan: e, g, n. Jennifer Pan: e, f, g. Jaime Settle: c, e, g, j. Natalie Jomini Stroud: d, e, g, l, n. Emily Thorson: g. Rebekah Tromble: e, g, l. Joshua A. Tucker: e, f, g, n. Magdalena Wojcieszak: e, g, n.

S1.3 Ethical Considerations

Researchers involved in the project considered a number of ethical concerns related to the research and designed the studies to minimize potential harms to the respondents involved in them, as well as any broader social harms.

Unlike other studies in the project, which required informed consent, the study reported here involved analyzing platform data for *all* adult US users but only in aggregate form and applying privacy protection rules (i.e., we only analyze the diffusion structures of posts shared at least 100 times by adult US users; there is no personally identifiable information in these data structures, as we explain in more detail in section S2).

Meta sought review from and was granted approval to conduct the experimental studies by the NORC Institutional Review Board (Protocol number 20.08.10, Project number 8870). Academic collaborators worked with their respective university IRB's to ensure compliance with Human Subjects Research regulations in their authorship of papers, including analysis of aggregated, de-identified data collected by Meta and NORC.

S1.4 Professional Ethics Advice

Meta retained the services of *Ethical Resolve*, a data ethics firm that was consulted by both Meta and academic researchers at various stages of the project prior to implementation of the research to evaluate whether it met long-running traditions of research ethics as well as emerging norms and best practices for conducting digital research.[S13]

S2 Study Design for this Paper

In section S1, we introduced the scope and characteristics of the *US 2020 Facebook and Instagram Election Study*. This paper is one of the multiple studies that, as discussed, were preregistered as part of the larger project. In this section, we offer details of the specific data that was used to produce the results in this paper. All figures reported in the main text, and most of the results discussed are based on aggregated platform data tracking *all* diffusion events generated by US-based users, Pages, and Groups between July 1 2020 and February 1 2021. In other words, the results we report are *not* based on recruited participants but on *all* adult US users that were active on the platform. This is the reason why we aggregate the data, as explained in subsection S2.1: we wanted to preserve the privacy of users who did not opt into the *US 2020 Facebook and Instagram Election Study* (which are the majority of Facebook users).

A small set of our analyses make use of the panel data collected by NORC to track on-platform behavior and attitudes at the user level and for a sample of participants who consented to providing this data (as explained in section S1 and discussed in more detail in subsection S2.2). In particular, we use this panel data for two main purposes: (1) to validate on-platform predictions of user ideology (as we explain in section S3.3, we compare the classifier predictions with the self-reported ideology of recruited panelists); and (2) to estimate how concentrated posting, sharing, and exposures to content are among users (i.e., a minority of users generate the majority of re-shares and we quantify the size of this minority using the individual-level behavioral traces of recruited panelists; individual-level data are necessary to produce these metrics of engagement, which we report in Figures S42 and S43). In S2.2 we offer more details

[S13] https://ethicalresolve.com/

of the panel data and how it was collected. In particular, we discuss information on sampling; participant recruitment and consent; response rates; and weights.

S2.1 Aggregated Platform Data

This study was designed to identify the types of content that are more likely to go viral on Facebook. The basic unit of analysis are diffusion trees. A diffusion tree is a branching structure where nodes are "actors" posting content, and branches encode the sharing activity of that content. Diffusion trees reconstruct the trajectory of information cascades and measure the reach of diffusion events (the terms 'trees', 'cascades', and 'diffusion events' are used interchangeably in this paper, within the context of the technical definitions below). This study compares the structural properties of those trees for (1) different types of content, created by (2) different types of entities (i.e. users, Pages or Groups), and shared by (3) users with different demographic characteristics. Ultimately, the goal is to identify differences in how content spread or diffused on the platform prior, during, and after the 2020 US Election.

In our data, every tree corresponds to a piece of content independently introduced on the platform via posts. This means that there are as many diffusion events as independently introduced posts (i.e., the same piece of content, e.g., text, URLs, pictures, videos, can be posted by multiple users but they count as independent cascades if they are not part of the same diffusion event or tree). In this section, we offer details on how we aggregated platform data in the form of these diffusion trees. The platform data includes all posts (original shares and re-shares) generated by all adult US-based active Facebook users between July 1 2020 and February 1 2021 (we refer to these data as 'MAP data', where MAP stands for Monthly Active People).

In intentionally building privacy protections into the research design, the academic research team did not have access to individual-level user data for all US adults. Rather, MAP data were shared with the academic team in a summarized, aggregate form. Code to collect and process the data from Meta servers was written by Meta researchers and then reviewed by at least one member of the academic research team. The academic researchers only had access to code that was implemented within Meta specifically for this project (e.g., pre-existing code to predict the political ideology of US users or to classify posts as civic/non-civic was not provided for review).

While the use of this MAP data is permissible under Meta's Data Policy, care was nevertheless taken to ensure that user privacy was protected. These procedures included the following:

- First, all requests were reviewed by Meta Legal and Privacy to ensure that no individual-level or identifying data were shared with the academic research team.
- Second, techniques were used to reduce the risk of any individually identifying data being shared. These techniques included the $k >= 100$ shares threshold applied to diffusion trees (i.e., trees smaller than size 100 are not part of our main analyses); replacing unique IDs for each post with randomly generated alphanumeric strings; reporting timestamps

with only hourly precision; and rounding proportions to two decimal numbers in order to reduce the probability that individual posts in the dataset can be identified.

S2.2 Panel Data

The analyses comparing self-reported and predicted ideology and measuring the concentration of misinformation shares rely on data from a set of participants who consented to provide their individual-level Facebook log data as part of this study. This sample was initially recruited for the purposes of studying how changes in platform features affect users attitudes and behavior through a series of experiments (whose results are reported in other papers within the broader project), but it is used here to evaluate the ideology predictions and to offer additional context on our results. To avoid the possibility that any impact of these interventions on respondents' sharing patterns may affect our conclusions, in this study we only rely on the control group of this sample, i.e., users not affected by the experimental interventions applied as part of other studies. Figure S2 summarizes the selection of panel users and participant flow through each stage of the selection process. Participants generating the data we analyze in this study were paid $5 for completing the surveys (waves 1 and 2, see below). The total sum of incentives paid to participants whose data is used in the analysis reported in Figures S42 and S43 is $97,790. The rest of this section describes in more detail the recruitment process and completion rates.

Sampling The sampling approach was designed to achieve specific sample targets across different stages of a six-wave survey panel (this panel data is more extensively used in other papers within the larger project). The sample targets for this study are provided in Table S1. These targets were chosen to achieve desired minimum detectable effect sizes (MDEs) for the project's experimental interventions (reported in separate papers) across different subgroups and among the set of respondents participating in wave 1 (recruitment) and wave 2 (baseline) surveys in addition to at least one of waves 4 or 5 of the study (we refer to these participants as *three-wave completes*). We worked backward from our target MDEs in waves 4 and 5 to determine the number of respondents we would need at the recruitment stage (initial 3-wave). The three-wave completes target sample for on-platform interventions was adjusted after observing the level of attrition between the wave 1 and wave 2. In response, and prior to treatment randomization, the academic researchers in the study decided to reduce the set of planned platform interventions, which also reduced the target sample size. The revised targets are provided in the last column of Table S1 (final 3-wave).

Sampling Frames and Stratification The sampling frames included all Facebook monthly active US-based users 18 years of age or older eligible to receive general surveys as of August 17, 2020. Participants were asked to confirm they were over 18 years of age and lived in the United States as part of the recruitment process. The Facebook sampling frame was trimmed by removing predicted fake accounts, employees, and advertisers.

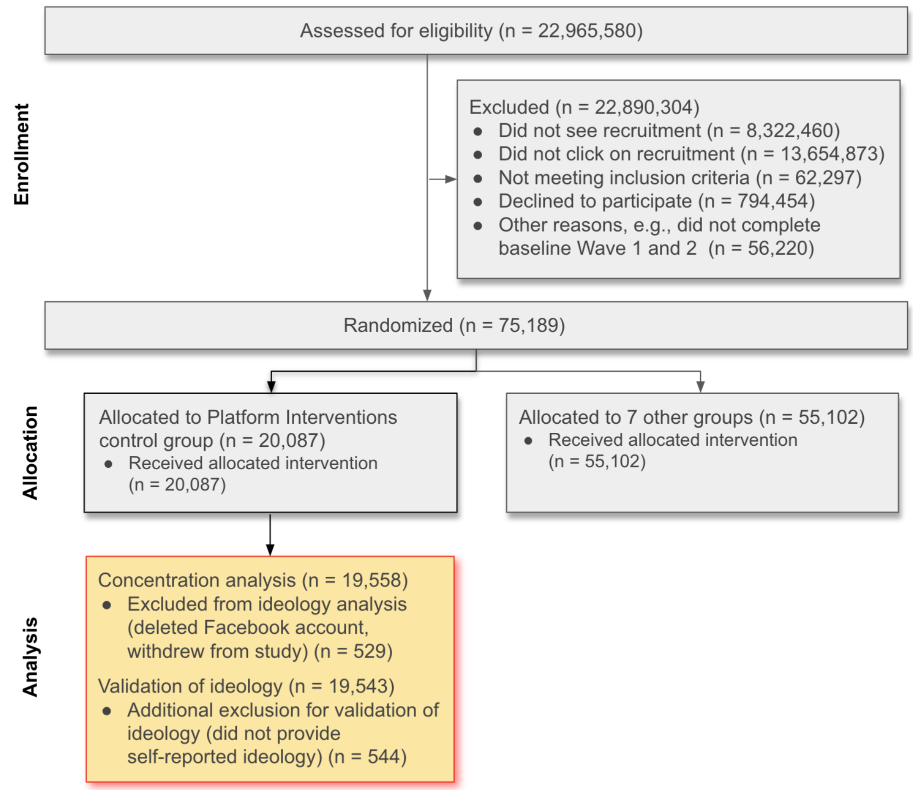

Figure S2: **Recruitment Flow Diagram.** A total of 75,189 users agreed to participate in the platform interventions conducted as part of the *US 2020 Facebook and Instagram Election Study* and were eligible (e.g., completed baseline surveys in Waves 1 and 2). Of these, 20,087 were randomly allocated to the control group, i.e., this group did not see any experimental intervention and therefore their user experience did not change. Our analyses comparing self-reported and predicted ideology and measuring the concentration of misinformation are based on this group of recruited participants. These are the data we use to produce Figures S42 and S43.

Table S1: Target size for the sample used in this study

		Sample targets		
Apps	**Study sample**	**Wave 1**	**Initial 3-wave**	**Final 3-wave**
FB	Platform interventions	119,000	71,118	54,740

The three-wave sample targets were revised after observing the wave 1 and wave 2 survey completions and adjusting the proposed number of treatment arms prior to randomizing participants into condition.

The sampling frames were stratified along several covariates of interest, i.e., number of days a user was active on the platform (the number of days a user logged in to Facebook in the 30 days on or before August 17, 2020, classified into three categories: 1–14 days, 15–29 days, and 30 days); a user's predicted census region (East, Midwest, South, West)[S14]; whether the user is predicted to live in a battleground state[S15]; a user's predicted ideology (liberal, moderate, or conservative)[S16]; and the census ethnic/racial composition in the zip code in which a user is predicted to live (percent of Hispanic residents and Black residents)[S17]. The stratification of the sampling frame for this sample generated 621 population cells.

Sampling Probabilities and Target Distributions Sampling probabilities were computed to achieve specific sample distributions for the set of demographics encoded in the stratification step across each of the samples of interest. The sampling probabilities took into account (a) differential non-response across different demographics[S18] and (b) the desired sample size across the different studies.

The Wave 1 target sample size took into consideration a total attrition rate of 40% between Wave 1 and the combination of Wave 2 and at least one of Waves 4 or 5 (three-wave completes) of the study while ensuring our ability to detect an MDE of 1.5 percentage points on turnout

[S14]The classification of states across each of the census regions is available here

[S15]Following the two most recent Electoral College Ratings by the Cook Political Report prior to August, we defined as battleground states those whose complete electoral geography was considered in the "Toss Up", "Lean Democrat", or "Lean Republican" in at least one of the reports. "Toss Up" states included: Arizona, Georgia, Maine, North Carolina; "Lean Democrat" or "Lean Republican" states included: Florida, Michigan, Minnesota, New Hampshire, Pennsylvania, Wisconsin, Iowa, Ohio, and Texas. Nebraska was excluded because only one of three congressional districts was identified as a battleground district.

[S16]See SM Section S3.3 for additional details on how ideology is predicted. Liberal users have predicted ideology score below 0.35, conservative users have a predicted ideology score above 0.65; moderates have a predicted ideology score between 0.35 and 0.65.

[S17]Some fields had missing values (e.g., predicted ideology, state, and zip code). Individual values were imputed probabilistically using the distribution of demographics in the population. In general, the percent of missing values for a given demographic was quite small, never exceeding more than a few percentage points of the population.

[S18]Responses to Facebook surveys with a similar design were used to model differential response rates.

Table S2: Target sample demographic distribution of survey sample

Demographic	Target Distribution
Number of days user logged in to app	Less than 15 days (4%), between 15 and 29 days (24%), and 30 days (72%)
Minority users (Black or Hispanic)	56% in battleground states and 58% in non-battlegrounds states.
Users in battleground states	40%
Predicted ideology	Conservative, liberal, moderate. No initial target.

and vote choice for the experiments (reported in separate papers). The specific distributions we aimed for the sample used in this paper are included in Table S2.

As shown in Table S2, the target distribution included 4% of those logging on for fewer than 15 days, 24% between 15 and 29 days, and 72% for 30 of the past 30 days. This includes an oversample of those using the platform more frequently. In battleground states, the target sample of minority (Black or Hispanic) users was 56%. In non-battleground states, the target sample was 58% minority users. The remaining rows show the targets based on geography and predicted ideology.

In meeting these targets, it is important to take into account three considerations:

1. All else equal, the probability of a user being invited to participate in the study in a given strata is proportional to its size in the sampling frame.

2. There was no initial target distribution for ideology. We incorporated this dimension to the stratification in the second week of recruitment after seeing that self-reported white liberal users were more likely to consent to participate in the study. We did so by oversampling, based on their predicted ideology, moderate and conservative users. No specific targets were identified, but the proportion of users who self-identified as Democrats was reduced.

3. Meta informed the academics that it had limited race/ethnicity data coverage for its US users and it could not be used to inform large scale probability sampling. Instead the probability that a given survey respondent identifies with a given ethnic/racial category was derived based on the ethnic/racial distribution of a user's predicted zip code. The implementation of this approach is presented in more detail in the 'Sampling for Race and Ethnicity' subsection below. As the approach to sample minority users had mixed results in the early stages of the recruitment period, and in light of observed imbalances in ideology, ideology targets were prioritized.

Implementation of Sampling Scheme After defining the sampling probabilities, the sampling scheme was executed as planned. In the implementation of the sampling scheme two additional steps were taken:

1. The sampling probabilities were adjusted when the size of a given stratum was exhausted. The adjustment to the sampling probabilities for non-exhausted cells across the sampling frame was done in proportion to their size.

2. Sampling was executed sequentially to avoid users being invited to more than one experimental intervention within a given app. This left a small probability that users of Facebook and Instagram could have been invited to participate in a similar or different experience across the two apps. This issue does not affect this study since we only rely on the sample recruited on Facebook.

Sampling for Race and Ethnicity The project aimed for specific sample proportions of Black and Hispanic users in the sample. Meta informed the academic team that Facebook and Instagram do not have data on the race/ethnicity of US users that could have informed our sampling decisions. Therefore, the probability a given user identified with one of the categories of interest within a given strata of the sampling frame was modeled. The focus was on achieving targets for the White and Other ethnic categories, which allowed us to obtain the desired distribution for minorities as the residual categories, based on people's predicted zip codes. The implementation of this approach involved the following steps:

1. Let i represent one of the stratum of the S strata in the sampling frame of a given app. Let X_i represent the vector of covariates defining the characteristics of strata i. The vector X_i has the following components: number of days users were active on the platform in the last 30 days, census region, battleground state, predicted ideology, zip code percent Black, zip code percent Hispanic.

2. Within a given X_i we compute the probability that a user belongs to one of three different categories: Black, Hispanic, White or Other. The following steps were followed:

   (a) We use the quartiles of percent Black and Hispanic in a given zip code. For these two categories we have four potential values 0–25%, 25–50%, 50–75%, 75%+.

   (b) The lower and upper bounds of the residual category are backed out of the combination of these quartiles. For example, in a zip code with 0–25% Hispanic and 0–25% Black, we know that the White or Other population represents at least 50% or at most 100%.

   (c) Based on this information, the probability of a user's race/ethnicity was derived using the midpoints of each of the quartiles across all ethnic categories. In the example introduced in the previous bullet, the probabilities were as follows: Pr(Hispanic) = Pr(Black) = 0.125 and Pr(White or Other) = 0.75.

(d) The only exception is when the bounds for White or Other are 0–25%. In this situation, we set Pr(White or Other) = 0.125 and define Pr(Hispanic) = Hispanic midpoint * (1 - Pr(White or Other)) and Pr(Black) = Black midpoint * (1 - Pr(White or Other)). (This solution applies to cases when the percent Hispanic is 0–25% and the percent Black is 75–100%.) Otherwise, if we simply used midpoints to assign probabilities would yield Pr(Hispanic) = 0.125, Pr(Black) = 0.875, and Pr(White or Other) = 0.125, which would sum to more than one.

3. For a given attrition and click-through rate we back out the number of wave 1 respondents in the White or Other category needed to achieve the target for this group in the three-wave complete sample. In particular, the number of wave 1 respondents in the White or Other category is:

$$n_o|stratum_i = pr(\textit{White or Other}|stratum = i) * pr(s = 1|stratum = i) * N_i$$

where $pr(s = 1|stratum = i)$ represents the probability a user in stratum i is invited to participate in the study and N_i denotes stratum i's population size.

Solving for the sampling probability in stratum i yields:

$$pr(s = 1|stratum = i) = n_o|stratum_i/(pr(\textit{White or Other}|stratum = i) * N_i)$$

4. Note that the $pr(s = 1|stratum = i)$ derived in the previous step implies that we have a total number of invitations of Black and Hispanic users. These are given by:

$$n_{Black}|stratum_i = pr(Black|stratum = i)) * pr(s = 1|stratum = i) * N_i$$

$$n_{Hispanic}|stratum_i = pr(Hispanic|stratum = i)) * pr(s = 1|stratum = i) * N_i$$

From these quantities we can back out the implied numbers of respondents in the wave 1 and three-wave sample in a given stratum.

5. We repeated steps 1–4 to obtain across all strata S of the sampling frame of a given app to try to obtain the desired targets. However, as noted above, we decided to prioritize over-sampling users on predicted ideology because (a) our initial approach exhibited mixed results in terms of meeting the specified minority targets, and because (b) the early re-cruitment numbers showed a skew towards Democrats. To correct the partisanship skew, we simply repeated steps 1–4 across each of the three predicted ideology strata (L, M, and C) with a higher proportion of survey invitations allocated to the M and C strata.

Figure S3: Image shown to recruit participants on Facebook

Participant Recruitment and Consent At the top of their Facebook feed, randomly selected participants saw a recruitment message asking them if they would like to share their opinion as shown in Figure S3. Those clicking "Start Survey" were directed to a consent form. Participants gave their consent to participate in the on-platform experiments using an IRB-approved consent form, as follows:

Do You Want to Participate in a Research Study About the US Election in November? *Your participation in this research will help researchers at New York University, The University of Texas at Austin, and other academic institutions, as well as Facebook, understand more about how people's experience with Facebook and Instagram affects their opinions and behaviors on elections.*

How it Works *Over the next four months, you'll be asked to fill out a short survey each month. This monthly survey will take about 15 minutes, for a total of 60 minutes over four months. Our partner, NORC at the University of Chicago, will administer this research. During this time, your [Facebook/Instagram] experience may be different than what you're used to. For example, you might:*

- *See more or fewer ads in specific categories such as retail, entertainment, or politics*
- *See more or fewer posts in [Feed / your feed] related to specific topics*
- *See more content from some [friends / connections] and less content from other [friends / connections]*
- *See more or less content about voting and elections*

You'll be paid at least $30 for participating in this study and completing all four surveys, including $5 for each of the first two surveys and $10 for each of the final two surveys.

- *You will receive your reward as an electronic gift card, delivered within 1 day of completing each survey*

- *You can only take each survey once*
- *If you do not complete the first survey, you will be removed from this study*

If you choose to participate in this study, your survey responses will be linked with your Facebook and Instagram activity data from the 2020 calendar year.

Benefits, Alternatives, and Risks *There are no benefits to participating in this research, nor are there risks greater than those encountered in everyday life, including risks related to the loss of confidentiality. You can learn more about how we're keeping your information safe in the Data Collection and Your Privacy section below. You can choose not to participate in this study.*

Data Collection and Privacy *If you choose to participate in the study, the following will happen:*

- *NORC will join your survey responses to publicly available third-party data like if you've voted or made a political contribution, if this data is available*
- *Facebook will combine this data with your activity on Facebook and Instagram from the 2020 calendar year, collectively called Combined Data*
- *This Combined Data will only be used for research purposes and will not be used to show you ads*
- *This Combined Data will be shared with our academic partners and, if legally required, with the Institutional Review Board (IRB) that reviewed this study*
- *All access to this Combined Data will be monitored and logged*
- *Once this study is over, de-identified data (i.e. data where identifiers such as your name and other information that could reasonably be linked to you are removed) will be stored and shared for future research on elections, to validate the findings of this study, or if required by law for an IRB inquiry*

You can decide to stop participating in this study at any time, for any reason, and without consequences. You may withdraw by visiting the study website hosted by our survey administrator, NORC at the University of Chicago, at 2020erp.norc.org. If you have any questions related to this research, you can email NORC at erpStudy@norc.org, or call toll-free at (866) 270-2602 between 9:00 AM - 10:00 PM ET.

If you are a research participant and have questions about your rights, or have concerns or complaints about this research, you can email the NORC Institutional Review Board (IRB) at surveyhelp@norc.org or call (866) 856 - 6672 between 9:00 AM and 10:00 PM ET. Please note *that by contacting or providing information to NORC IRB, NORC IRB may obtain information about you, including any personal information that you share. Even though NORC IRB is affiliated with Facebook as this research study's IRB, Facebook's Data Policy does not apply to any information about you shared with NORC IRB when you initiate contact.*

If you join this study, you affirm that you are at least 18 years of age and live in the United States. Once you join this study, you'll be sent off [Facebook/Instagram] to a site hosted by our study administrator, NORC, to complete a 5-minute enrollment form.

Data Collection Timeline Data collection began with a soft launch on August 31, 2020 and continued through March 2, 2021. In this study we only use data from the first two survey waves, which were fielded as follows:

- Wave 1: A subsample of Facebook-recruited respondents were invited to the survey on August 31 in a soft-launch. The remainder of sampled Facebook/-recruited respondents were invited to the survey on September 1. The recruitment of the sample continued until Saturday, September 12. The wave included the recruitment and consent processes and a short survey.
- Wave 2: The field period for Wave 2 started on September 8 and continued through September 23. The wave included a baseline survey and was conducted prior to randomization into a control group and a series of different platform intervention treatment groups. As noted above, in this study we only use data from the control group.

Response Rates Table S3 contains details on the survey response rates. Eligibility was determined based on respondents' age (above 18) and removing duplicate cases. While the study continued for up to six survey waves and contained other samples and platforms, here we only report the rates that are relevant to the data we use in our study: the Facebook sample recruited for on-platform interventions that completed waves 1 and 2 of the survey. The control group that we analyze in this study corresponds to a randomly selected subset of 35% of participants within this sample.

Table S3: Survey completion and response rates

Row	Measure	FB Intervention sample	Definition / Formula
	Pre-NORC Recruitment		
A	QP eligible users sampled	22,965,580	
B	QP viewers	14,643,120	
C	QP clickers	988,247	
D	DE willing	n/a	
E	DE amount selected	n/a	
F	Consented to full study	193,880	

G	Did not consent to full study	794,367	C-F
H	Subsample of study non-consenters asked to consent to surveys only	n/a	
I	Consented to surveys only	n/a	
J	QP viewers among sampled	63.8%	B/A
K	QP clickers among QP viewers	6.7%	C/B
L	DE willing among QP clickers	n/a	
M	DE amount selected among DE willing	n/a	
N	Consented among asked	19.6%	F/C
O	Response rate before pass to NORC	0.8%	J*K*N
	Wave 1		
P	Passed to NORC	189,792	
Q	Screened for eligibility	163,207	
R	Confirmed eligible	162,698	
S	Completed wave 1	139,193	
T	Screener completion rate among passed to NORC	86.0%	Q/P
U	Eligibility rate among screened	99.7%	R/Q
V	Interview completion rate among eligible	85.6%	S/R
W	Response rate among those passed to NORC	73.6%	T*V
X	Cumulative response rate, pre-NORC recruitment wave 1	0.6%	O*W
	Wave 2		
X	Invited	139,193	
Y	Screened for eligibility	77,438	
Z	Confirmed eligible	77,405	
AA	Completed wave 2	75,318	
AB	Screener completion rate among invited	55.6%	Y/X
AC	Eligibility rate among screened	100.0%	Z/Y
AD	Interview completion rate among eligible	97.3%	AA/Z
AE	Response rate among invited, wave 2	54.1%	AB*AD
AF	Cumulative response rate, pre-NORC recruitment waves 1 2	0.3%	X*AE

Weighting for Panel Data For the on-platform interventions sample we use in this study, survey weights were created to generalize treatment effects to the best estimate of adult monthly active users: all US Facebook monthly active users 18 years of age or older eligible to receive general surveys as of August 17, 2020. The general approach to creating the weights was to reduce bias while maintaining a low design effect.

Inverse Propensity Scores Weights (IPSW) using LASSO regression with Facebook log data were built. Covariates used for block randomization in the experiments and variables presumed to predict treatment heterogeneity were prioritized. The weights calibrate to the full population of Facebook users and they were built using:

- Predicted ideology (divided into liberal, moderate, and conservative using the ideology classifier described in SM S3.3 with cut points of 0.35 and 0.65).
- Friend count, divided based on terciles.

- Political Pages followed, divided by tercile.
- The number of days a user logged on to their account in the 30 days prior to August 17, 2020, divided into 29 or less vs. 30.
- Whether the respondent was in a swing state.

We used raking to create the set of final weights that calibrate to population estimates of race (white vs. non-white), party ID (Democrat, Independent, or Republican, including leaners as partisans), and education (less than a college degree vs. a college degree or more). The specific targets are based on the distribution of these characterstics among self-reported Facebook users in a separate dataset (the AmeriSpeak panel) that is representative of the US adult population.

Our final step was to trim the weights. Following the Cooperative Election Study[S19], which trims weights above a particular threshold, and the Pew Research Center[S20], which has trimmed weights at the 1st and 99th percentiles, we trimmed the top 1% of the survey weights.

We did not include design weights in the computation of the survey weights as the weights increase the design effect significantly without appreciably decreasing the bias.

Coding of On-Platform Behavior For the set of participants who consented to join our panel and provide individual-level Facebook log data, we collected a series of metrics based on their on-platform behavior. In this paper, we focus on six of them, which relate to their production of content. In particular, we measure the count of new posts created by respondents between September 24th, 2020 and December 22, 2022, which correspond to the dates for which the platform interventions were running. These counts are disaggregated by whether they are original posts or re-shares of other existing posts; and whether they are predicted to be political or misinformation. (For additional details on post-level classifications, see Section S3.3.) All types of posts are considered here (text posts, url posts, image posts, photo albums, etc.); and we include posts created on any surface (e.g. on their profile, on NewsFeed, as a Group post, etc.).

S3 Materials and Methods

Our analyses rely on the reconstruction of information cascades and diffusion pathways in the form of network trees. These trees (and their structural properties) are our main unit of analysis. We also analyze the composition of the trees in terms of user and message characteristics. The following two sections explain in more detail the logic behind these data aggregations and the variables we use to summarize the data.

Our main dataset (large trees or posts created by US-based adult users that received at least 100 re-shares) comprises 12.1 million Facebook posts that were re-shared by 114.3 million

[S19] https://sda.berkeley.edu/sdaweb/docs/cces2018/DOC/CCES+Guide+2018.pdf.
[S20] https://www.pewresearch.org/methods/2021/04/08/appendix-a-standard-atp-weighting.

unique adult US-based users and 2.1 million Pages (including re-shares into 4.3 million unique Groups) approximately 11.1 billion times in total. In section S5 we discuss additional results that complement those discussed in the main paper. In section S6 we also report some aggregated statistics for *all* posts created by US adult users, including small trees (which amount to 98.8% of all trees with at least 1 re-share).

S3.1 Diffusion Tree Data

Tree data structures are hierarchical networks with a root node on level 0 and nested layers of additional levels created by parent/child nodes: every parent node can have several children, but every child node can only have one parent node. Nodes with no children are called 'leaves', and they are endpoints in the tree. In the context of our data, nodes are posts published on Facebook (in public or private mode); and edges indicate re-sharing behavior. Figure S4 offers a schematic representation of some simple tree structures that can emerge during the diffusion of information.

We summarize the properties of these trees using three basic statistics: size or number of nodes (i.e., re-posts of the original root message); depth or maximum distance from a leaf node to the root; and breadth or maximum number of nodes over all depths or levels. The example tree in panel C of Figure S4 has size 9, depth 3, and maximum breadth 4 (on level 2). Following prior work (*1, 2*), we also calculate a fourth statistic, called structural virality, which calculates the average distance d between all pairs of nodes in a diffusion tree T (with $n > 1$ nodes) according to this formula:

$$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}, \tag{S1}$$

The distance d_{ij} measures the length of the shortest path between any pair of nodes. For instance, the distance between nodes i and j in panel C of Figure S4 is 2. Averaged over all pairs of nodes as indicated in equation S1, the structural virality of this tree is 2.82. To give another example, the same statistic for tree 2 in panel B is 1.67 (the size for this tree is 5, the depth 1, and the maximum breadth 5). Table S4 displays additional information about the distribution of these metrics in the dataset that we analyze.

Ghost Branches Our data encompasses all posts that were published or re-shared by all adult US-based active Facebook users. This means that posting and re-sharing activity generated by non-US users or users younger than 18 years old does not appear in our data, a data gap that manifests in the form of missing intermediary nodes for about $\sim 49\%$ of all the trees we analyze. (For trees where the original post is not US-based, we do not have a root node; we dropped these trees, about 3% of all trees, from the data). Our data keeps track of where ghost branches start in terms of depth because this metric is logged for each post in Facebook's data tables, but we do not have the data to reconstruct the specific path by which these branches

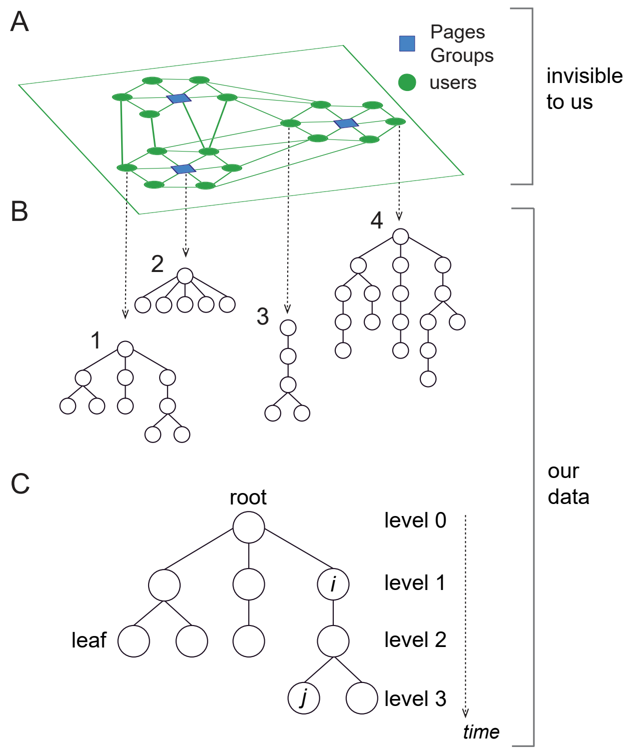

Figure S4: **Schematic of our Data.** This Figure illustrates the structure of our data. Our unit of analysis are the diffusion trees that emerge through the sharing of posts (public and private). These trees grow in the underlying social graph, but the structure of that graph or the network position of the tree initiators (i.e., users/Pages/Groups posting the original message that is then re-shared) are beyond the scope of our data. However, the nested layers in the diffusion trees are shaped by the social graph (i.e., users in level 1 are more likely to be connected to the creator of the root post; users in level 2 are more likely to be two steps removed, etc). We summarize the structure of the diffusion trees with the statistics discussed in the text (i.e., size, depth, breadth, and structural virality). The example tree in panel C has size 9 (the original post is not counted), depth 3, and maximum breadth 4. Tree number 2 in panel B has size 5, depth 1, breadth 5, and structural virality 1.67.

were originally attached to the tree because the intermediate nodes are non-US users. In Figure S5 we illustrate the different options we had to reconnect these trees. We opted for the option displayed in panel C (minimal new connection). It is important to note that the presence of these ghost branches only affects the calculation of average distances and, therefore, only the measure of structural virality. Trees without ghost branches have a median virality of 2.83; trees with ghost branches have a median virality of 3.35.

Attribution Rule When building the cascades, we rely on an attribution rule that parses the re-sharing behavior of users, i.e., we create a tie connecting a node (a post) with its most immediate parent node (the post that triggered the re-share) every time users explicitly click on the share button (which Facebook introduced in 2010). The main advantage of this approach is that it allows us to track an explicit diffusion mechanism without inference error, and it relies on the same approach as other studies of re-share cascades on Facebook (e.g. *3*). This is also the rule that other research with Twitter data tries to approximate when reconstructing their diffusion trees: in (*2*), for instance, they use time-inferred diffusion cascades because their data did not allow them to observe directly who clicked on whose post's re-share button. This data limitation forced the researchers to infer child-to-parent ties using timestamps and contrasting the network of followers to determine who was the parent node. Our approach has less measurement error than this Twitter work since we don't infer re-shares, we observe them; and it is directly comparable to past research that also analyzes diffusion chains on Facebook (again, at least, since the share button was introduced). However, it is important to note that other attribution rules are possible to account for behaviors that do not involve clicking on the re-share button[S21].

For instance, users may copy and paste content instead of click on the share button. In this case, we would assume the post is the root of a new tree, thus missing a diffusion tie in an existing tree (and thus misrepresenting its size). Users may also click on the share button on the content published by the root node as opposed to the parent responsible for their exposure. This happens, for example, when we see something from a friend but click on the share button of the original source that our friend re-shared instead of on the button of the friend's post. To illustrate it further: if user A publishes a post, and user B re-shares it, user C sees B's post but may decide to re-share A directly, bypassing B (even if B is the reason C saw that content); user C may also only be able to re-share A if user B's post is set as 'private' (in which case, they wouldn't have much of a choice as to who to re-tweet). These scenarios would result in a tree with the structure: A $\rightarrow$ B, A $\rightarrow$ C instead of the actual causal structure A $\rightarrow$ B $\rightarrow$ C – causal in terms of capturing the actual information flow or flow of influence. In this case, the size of the tree would still be accurate but its structure would be misrepresented.

We do not argue that our trees are exact representations of all the causal pathways that allow people to become aware of content. What we argue is that our trees are very accurate representations of re-sharing behavior, and re-shares are the main mechanism for information

[S21] We are grateful to Dean Eckles for raising this important point to our attention, and the considerations that follow.

diffusion on Facebook (it was designed for that purpose, and the purpose of amplifying reach). According to past research, assuming that all trees grow out of the attribution rule we apply to our data slightly misrepresents the depth of the resulting trees (see Figure 2 in *4*, although the dataset used in that paper is now more than 10 years old and it only includes posts with photos). What all these considerations highlight is that the choices made when reconstructing diffusion trees can impact, potentially, the statistics calculated from those trees. For the purposes of this paper, we decided to align with an attribution rule that is common in past studies and that minimizes inference errors – at least of re-sharing behavior that the 'share' button facilitates.

It is also worth noting here that cascades are not just the product of individual choices to post and click on re-share buttons, but also of algorithmic decisions to surface or demote content on users' Feeds (*5*), which may make some posts more likely to be re-shared. As we explain in the main text and detail in the timeline shown in Figure S11, our observation window includes periods of high-intensity content curation policies designed to curtail diffusion dynamics (these content curation policies go beyond what the Feed ranking algorithm is designed to accomplish on its own). This is why we focus on analyzing how the structure of diffusion trees changes across the different intervention periods: to assess the impact of temporally-bounded content curation policies.

Size Threshold As explained in the main text, and in sections S1.1 and S2, our analyses only consider trees with size $k => 100$ to further protect privacy (in what follows, we refer to this subset of trees as "large trees"). Figure 1A in the main text shows how many of the posts that were shared at least once get excluded from our analyses, given this threshold. During our observation period, there were $N \sim 1B$ posts published by US-based users (in public or private mode); only 12.1 M of these (about 1.2%) were re-shared 100 times or more. As we also show in Figure 1 in the main text, these trees (i.e., trees with size $k => 100$) accumulate most of the views (54.6%). In section S6 we offer additional aggregate analyses for the subset of trees below the size threshold.

Size-Matched Samples The measures we use to characterize trees (structural virality, depth, and breadth) are all correlated with the size of trees so, following prior research (*6*), our analyses include size-matched samples of trees – which allows us to run comparisons in tree distributions holding size constant. To do so, we created a size-matched corpus of trees for each comparison by considering each tree within the less frequent category (e.g., misinformation trees) and randomly sample a matching cascade from the other group(s) (e.g., political trees that are not misinformation), uniformly at random with replacement, from the set of trees of the same size. Cascades without a match of the same size (a rare occurrence given the size of our dataset; e.g. only 0.4% of misinformation trees cannot be matched) are excluded from the analysis of matched samples. The results we report in the main text of this paper are based on a single (randomly selected) matching step. In the SM, we report results based on repeating this sub-sampling 100 times in order to provide details on the distribution of test statistics obtained across realizations.

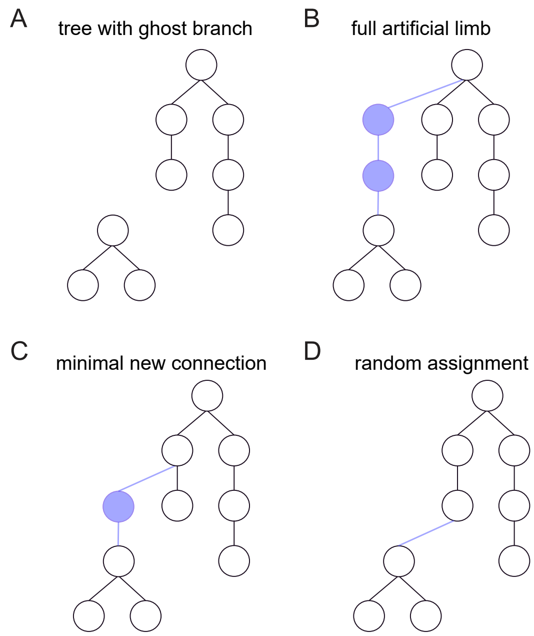

Figure S5: **Trees with Ghost Branches.** This Figure illustrates the different possibilities to reconnect a tree that has missing branches due to the boundaries of our data collection (only US-based Facebook users). We opted for option C, i.e., adding a minimal new connection. This is only relevant for the calculation of structural virality. About 49% of the trees have ghost branches. The median structural virality for trees without ghost branches is 2.83; for trees with ghost branches, it is 3.35.

S3.2 Tree Composition Data

In addition to structural properties, we also have attribute data characterizing the origin and composition of the trees. In particular, we look at (1) who published the root post (users, Pages, or Groups)[S22]; and (2) the tree classification as political/non-political; COVID-related; and/or misinformation. In this section we describe the logic of these additional tree variables; in sections S3.3 and S3.4 we discuss the technical details of the classifiers and measurements.

Political Trees Political trees were identified using Meta's Civic classifier (S3.3). This classifier operates at the post and re-share levels: every time a post is re-shared, the classifier makes a new designation, based on the additional information collected as more users interact with the post (in a sort of Bayesian updating process). The vast majority of trees do not see a change in the classification of the root post as it cascades, but we observed some change in the classification of larger trees. Our tree-level predictions of whether the content is political are based on a majority rule, i.e., we classified trees as political if 50% or more of the nodes in the tree were classified as civic. When compared to a classification of trees based on the root post only, the majority rule produces nearly identical results: for trees with 100+ shares, these methods disagree less than 4% of the time.

Political News Trees Political news trees were identified using Meta's Civic *and* News classifier (see section S3.3). As in the classification of all Political Trees, here we also apply a majority rule and classify trees as political news if 50% or more of the nodes in the tree were classified as civic and as news. In this case, we find that this approach disagrees only 1.1% of the time with an alternative method based on the categorization of the initial post in the tree.

COVID-related Trees COVID-related trees were identified using Meta's internal COVID classifier (S3.3). Again, this classifier operates at the post and re-share levels: every time a post is re-shared, the classifier may yield a different label (as it leverages information about the new users entering the cascade). We applied the majority rule here as well. In this case, compared to a classification of trees based on the root post only, the majority rule agrees $\sim 99\%$ of the times.

Misinformation Trees Posts are labeled as "misinformation" if either the post itself or any content that is part of it (i.e. a URL, a video or an image) is rated "false" by one of Meta's independent fact-checking partners[S23]. These independent, third-party fact-checkers are certified

[S22] We denote as user posts any posts by a user on their profile; and as Group posts any post by a user or a Page in a Group

[S23] See this fact-checking documentation, https://www.facebook.com/business/help, for a detailed description of rating categories.

by the non-partisan International Fact-Checking Network. The full list of partners (all independent to Meta) is available in the International Fact-Checking Network website[S24]. In the US, this list includes organizations such as Snopes, Reuters, The Washington Post, Fact Checker, FactCheck.org, and PolitiFact. All fact-checks are publicly available on the websites of these organizations and can be reviewed by any external source for accuracy. Since fact-checker ratings can change over time as new information emerges, we decided to use the list of ratings as of 2021-02-15, i.e., the list of what was considered misinformation at the end of our study period.

One important caveat in the identification of false content is that Facebook exempts politicians from their third-party fact-checking program. As stated on their Elections and Political Speech documentation[S25], Facebook did not send organic content or ads from politicians to their third-party fact-checking partners for review. However, according to this policy, "when a politician shares previously debunked content including links, videos and photos, we plan to demote that content, display related information from fact-checkers, and reject its inclusion in advertisements". Given the parameters of the collaboration, we (academic leading authors) were not given access to the list of Pages excluded from fact checking.

After content is fact-checked and labeled as misinformation, it is demoted in the Feed, it receives a warning screen, and it is no longer recommendable[S26]. However, some users may still see the content on their Feed, and find it if they search for it. In other words, the diffusion of misinformation does not necessarily stop after the content receives a rating from the external fact-checkers, but the added friction will slow down the diffusion. Our analyses looking at the growth of diffusion trees over time (S23 and S24) suggests that most of the re-shares happen within the first 24 hours. If fact-checkers take longer than 24 hours to rate content, their rating will not affect the reach of diffusion by much. It is also important to note here that, in addition to the Feed, there were other important content curation interventions during our observation period, including the "break the glass" measures discussed in the timeline (S11). These measures went above and beyond what the Feed algorithm usually does, creating different information regimes during the time window we analyze.

As we explain in the main text, $N \sim 114,000$ trees in our data ($N = 0.9\%$) are labelled as misinformation. In addition, in some of our analysis we compare posts in this category to posts that received other ratings by the independent fact-checkers, which we group into two categories: "misleading" (i.e., altered image, audio or video; partly false; or missing context) and "true" (i.e., posts that received an explicit "true" rating)[S27]. In addition to the misinformation trees listed above, approximately 89,000 trees in our data are labeled as misleading but only $N \sim 2200$ are labeled as true; the other $N \sim 12M$ are unrated, which means that they were not

[S24] https://ifcncodeofprinciples.poynter.org/signatories.

[S25] https://about.fb.com/news/2019/09/elections-and-political-speech.

[S26] See https://transparency.fb.com/enforcement/taking-action for more details on 'Penalties for sharing fact-checked content'.

[S27] Additional information about the meaning of these ratings can be found at https://www.facebook.com/business/help/341102040382165.

evaluated by Meta's independent fact-checking partners.

In general, we are most likely undercounting the total volume of false content circulating on the platform (i.e., posts that were rated by fact-checkers, or containing content matching other posts that were fact-checked, only represent a fraction of all posts). Prior research has also acknowledged this limitation, and the selection bias that may arise from the restriction of trees analyzed to only fact-checked content (*2, 6, 7*). However, it is important to note that, unlike this prior research, we use *all content* published on Facebook as the comparison benchmark, i.e., not just content that has been fact-checked and deemed true. This gives our data a much broader scope.

In Figure S6 we show changes in the count of all trees initiated on each day during our observation period, as well as counts for the subsets classified as political, misinformation, and COVID-related. Figure S7 shows the same information but expressed as proportions. On a given day, about 20-30% of all trees are political; $\sim 15\%$ to $\sim 5\%$ are COVID-related; and less than 2% are labelled as misinformation. Figure S8 shows the counts of trees initiated by users, Pages, and Groups. This figure confirms the message of figures 1F-H in the main text: while Page posts are the root source for most of the political trees, trees classified as misinformation are predominantly initiated by user posts (S8 panel C). This pattern, i.e., that Page-originated trees are generally larger but it is users who generate most large misinformation trees, is notable but not surprising. It is consistent with the deterrent effect of content moderation policies targeting Pages and Groups. The development of user-to-user sharing patterns evaded this enforcement policy. Note that we also find similar results for 'untrustworthy' sources (i.e., sources that had posted at least two pieces of misinformation since 2018 (see S3.4). Finally, we also see a drop in large tree counts right after the election, during the period of intensified platform interventions designed to curtail even more the spread of certain political content (see S11). Figures S9 and S10 offer more detail to these temporal trends.

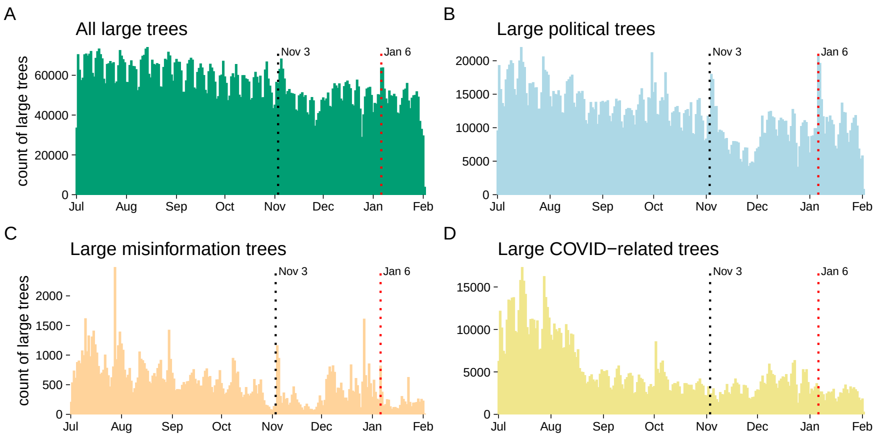

Figure S6: **Changes in the Volume of Large Trees (Size $k \geq 100$).** This figure shows changes in the count of (A) all trees; (B) trees classified as political; (C) trees classified as misinformation; and (D) trees classified as COVID-related that were initiated each day during our study period and accumulated a total of 100 re-shares or more. In addition to expected weekly seasonality, the decreasing trend in the number of all large trees and political trees could be related to a decrease in overall volume of posts (i.e., trees with size $k < 100$).

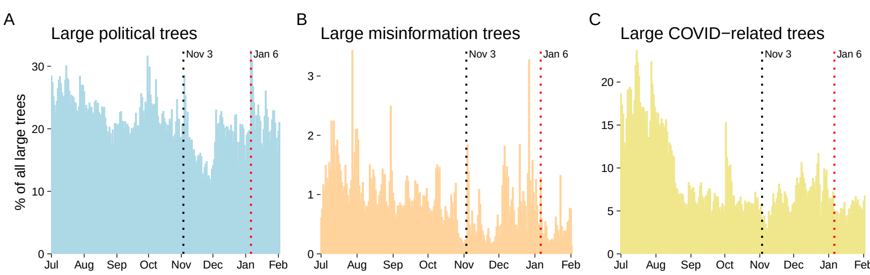

Figure S7: **Proportion of Large Trees.** This figure shows the same information as panels B-D in Figure S6 but expressed as the proportion of all large trees. About 20-25% of large trees are classified as political, less than 2% as misinformation, and $\sim 15\%$ to $\sim 5\%$ as COVID-related. There is a visible decline in the number of political trees of size $k => 100$ right after the election and a spike in trees flagged as misinformation a few days before January 6.

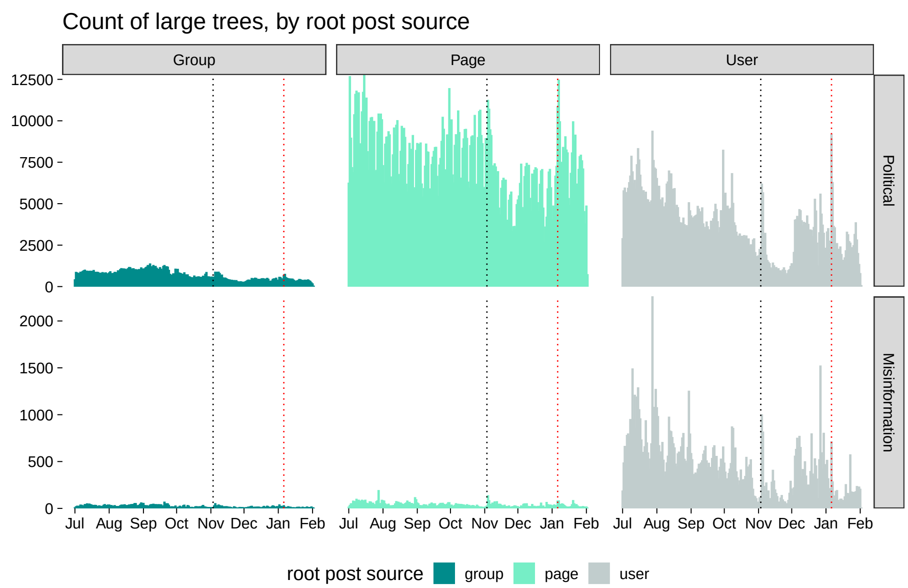

Figure S8: **Number of Large Trees Initiated by Users, Pages, and Groups.** This figure is an extension of panels F-H in Figure 1 in the main text, offering the counts of trees, not just the proportion. Pages overwhelmingly generate most large political trees, but users are clearly generating most of the large trees labeled as misinformation. We also see here evidence of a decline in the number of political trees right after the election; the numbers bounce back towards December.

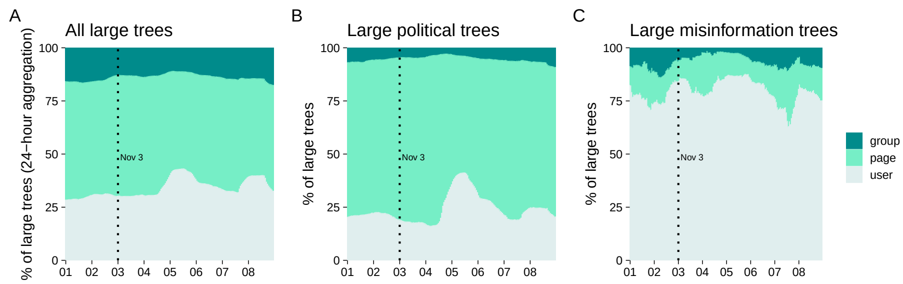

Figure S9: **Proportion of Large Trees Initiated by Users, Pages, and Groups around November 3.** This figure is an extension of panels F-H in Figure 1 in the main text, offering daily proportions based on rolling sums over the last 24 hours.

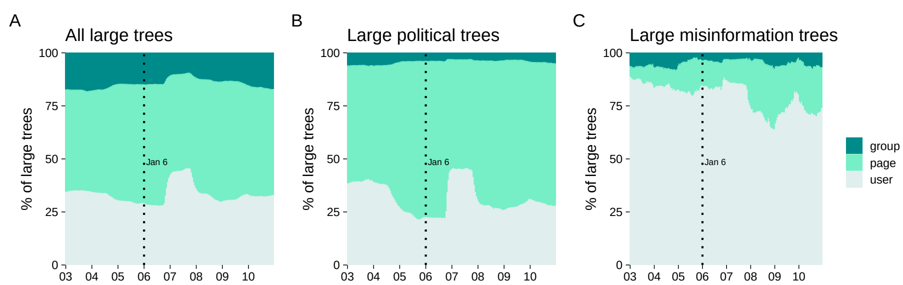

Figure S10: **Proportion of Large Trees Initiated by Users, Pages, and Groups around January 6.** This figure is an extension of panels F-H in Figure 1 in the main text, offering daily proportions based on rolling sums over the last 24 hours.

Other Variables In addition to content-based classifications, we characterized trees using other variables. In particular, we computed the average age for the users participating in each tree; the proportion that were classified as liberal/conservative (see S3.3); the proportion classified as having high-low political interest (see S3.4); whether the author of the root post was located in a swing state (S3.4); and whether the root post was 'boosted' as opposed to gathering organic (unpaid) views (S3.4). These variables are included in the analysis to allow us to characterize the populations involved in the dissemination of content on the platform, and whether content was seeded through paid posts. Since we do not have access to individual-level data (given our privacy protection measures, as we explain in S2), all these variables are aggregated at the tree level. Table S4 displays summary statistics for all these metrics.

Tree metrics					
Metric	p5	p50	p95	Avg	N
Tree size	105	215	2477	918.80	12086856
Tree depth	2	5	18	7.21	12086856
Tree max breadth	28	125	1404	425.98	12086856
Tree structural virality	2.05	3.05	10.30	4.28	12086856
User composition metrics					
Metric	p5	p50	p95	Avg	N
Average age	27	51	65	48.77	12068768
% liberal	0	24	93	32.12	12086856
% conservative	2	55	99	51.80	12086856
Ideological score	-0.96	0.38	1	0.19	12082702
% high interest (eng.)	15	63	95	61.02	12086856
% high interest (views)	5	35	79	38.02	12086856
% users in swing state	3	39	87	39.13	12086856
% reshares by Pages	0	0	2	0.46	12086856

Table S4: **Summary statistics (large trees).** This table reports summary statistics (5th, 50th, and 95th percentiles, average, and count of non-missing values) for all tree-level metrics and user composition variables.

S3.3 Meta Classifiers and Categorization Methods

The following classifiers, concepts, and categorization methods were either developed at Meta or are defined under Facebook and Instagram's platform policies.

Civic and News Classifiers We use Meta's internal classifiers to label trees as being about political content (Civic classifier) or about political news (Civic classifier *and* News classifier).

- Civic classifier
  - **Definition:** This classifier aims to capture content that relates to either *politics* (government, elections, politicians, activism, etc.) or *social issues* (major issues that affect a large group of people, such as the economy, inequality, racism, education, immigration, human rights, the environment, etc.). In this paper, we refer to any content that is classified as being in either of these two categories as "political"; otherwise they are not political.
  - **Usage:** We use the classifications for Facebook posts as applied to any post that was seen by US users during the US 2020 research project study periods. We use the classifications for both English- and Spanish-language content.
  - **Performance:**
    - Based on a sample of approximately 10k labeled posts, the classifier has 83% precision and 82% recall on English-language Facebook content.
    - Based on a sample of approximately 17k labeled posts, the classifier has 81% precision and 85% recall on Spanish-language Facebook content.
- News classifier
  - **Definition:** This binary classifier aims to capture content about current events, timely information, and that follows journalistic standards such as citing sources and having a byline[S28].
  - **Usage:** We use the classifier predictions for Facebook posts with a link that were created, seen or engaged with by US active users during the US 2020 research project study periods. We use the predictions for both English- and Spanish-language content.
  - **Performance:** Classification thresholds were chosen to yield 80% recall. Based on a sample of approximately 52k labeled links, precision at this threshold is 90% for US English-language News links. Based on a sample of approximately 36k labeled links, precision at this threshold is 42% for US Spanish-language News links.

[S28] See https://www.facebook.com/business/help/224099772719228 for more detailed information.

COVID Classifier We use Meta's internal COVID classifier to determine whether the trees are predicted to be related to the COVID pandemic.

- **Definition:** This classifier aims to capture whether a given post is related to the Coronavirus Disease 2019; i.e., content that explicitly mentions COVID-19; health concepts related to the virus; policy or public health guidance or the economic impact of the virus; as well as people's reactions to the pandemic.
- **Usage:** We use the classifier predictions for Facebook posts that were created, seen or engaged with by US users during the US 2020 research project study periods. We use the predictions for both English- and Spanish-language posts.
- **Performance:**
  - Based on a sample of approximately 30k labeled posts, 94% precision and recall for English-language Facebook posts.
  - Based on a sample of approximately 31k labeled posts, 94% precision and 95% recall for Spanish-language Facebook posts.

US Ideology Classifier The labels "liberal" and "conservative" are assigned to users based on Meta's internal US ideology classifier.

- **Definition:** This classifier aims to predict adult US active Facebook users' political ideology.
- **Methodology**
  - This classifier is trained to predict the ideology of US Facebook users based on their demographics, preferred language, location, and engagement with Pages. It outputs a numeric score ranging from 0 (indicating a user is predicted to be left- or liberal-leaning) to 1 (indicating a user is right- or conservative-leaning).
  - The classifier has high coverage — i.e., it is able to place up to 95% of adult US monthly active Facebook users on this numeric scale. The classifier also has high week-over-week stability -– on average, the weekly scores for individual users have a correlation of 0.96.
  - To further categorize adult US monthly active Facebook users into distinct ideological groups, we discretize the continuous ideology scores as follows: users with a score less than 0.35 are categorized as liberal. Those with a score greater than 0.65 are categorized as conservative. The remaining users with a score between 0.35 and 0.65 are categorized as moderate.
  - **Entity audience ideology scores:**

    * We use the user-level ideology scores to generate similar ideology scores for Pages, and Groups. We do so by computing a measure of the ideological composition of their audience: the average predicted ideology of Facebook users engaging with these entities in the last 28 days.
    * This approach places Pages and Groups on the same 0-1 numeric ideology scale as users. Since, by construction, these scores will converge towards the center given that they are averages of individual scores, here we select the thresholds of 0.40 and 0.60 to categorize the audience of each Page or Group as liberal (below 0.40) or conservative (above 0.60). For the middle category, We use the term "mixed" rather than "moderate", as the scores reflect the ideological composition of entities' audiences and not the ideology of the entities themselves.

- **Usage:**
  - We use the ideological classifications of US monthly active Facebook users 18 years or older who were active during the US 2020 election research project study periods.
  - We use the audience ideology classifications for all Pages and Groups producing content that US active Facebook users saw or interacted with during the US 2020 election research project study periods.
- **Performance:** We evaluated how well the user-level predictions performed at inferring self-reported ideology by comparing our classifications to the survey responses of US 2020 Facebook and Instagram Election Study panelists (see SM section S2). Precision and recall for each ideological group are as follows:
  - 3-class classification for users (using 0.35 and 0.65 as thresholds)
    * Self-reported Liberals: 62% precision and 77% recall
    * Self-reported Moderates: 53% precision and 30% recall
    * Self-reported Conservatives: 59% precision and 74% recall
  - We also evaluated how well the audience ideology measure might approximate the ideological affiliation of the entities themselves, if one exists. We compared the audience ideology scores we computed for the official Facebook Pages of US Members of Congress to a widely-used external measure of those Congress members' ideology derived from their legislative voting history, DW-NOMINATE, finding a 0.96 correlation (N = 409).

S3.4 Other Categorization Methods

The following categorization methods were either proposed by the academic team or adapted from published academic research.

Untrustworthy Trees In addition to labeling trees that contain misinformation (as explained in S3.2), we add an additional attribute indicating whether the source of the tree's root post is an untrustworthy Page or Group and/or whether the root post contains a URL from an untrustworthy domain. Pages or Groups are considered 'untrustworthy' if they have 2+ posts rated 'false' since the "misinformation repeat offender" program began in August of 2018; web domains are considered 'untrustworthy' if they have 2+ URLs rated 'false' by 3PFC since the "misinformation repeat offender" program began in August of 2018. In other words, the list of untrustworthy Pages, Groups, and domains is static and (as with misinformation) based on the ratings that existed on 2021-02-15 (the end of our study period). We use this measure in some of the analyses contained in the SM as an alternative way to identify potentially problematic content. There are two main reasons for this approach:

- The coverage of the 'untrustworthy' label is higher than 3PFC ratings: 0.9% of large trees are misinformation, according to 3PFC ratings, but 7.8% of large trees have a link to an 'untrustworthy' domain or have a root post whose source is an 'untrustworthy' Page or Group.
- 'Untrustworthy' trees are a superset of 'misinformation' trees: large trees with links to 'untrustworthy' domains are 10 times more likely to contain misinformation; and 62.1% of large misinformation URL trees correspond to trees with a link to an 'untrustworthy' domain. These numbers are similar for 'untrustworthy' Pages and Groups.

However, the 'untrustworthy' label does not apply to many of the posts users publish, which means that we lose significant data coverage when considering this attribute. Only user-generated trees that also contain a URL to a web domain can receive the 'untrustworthy' label. Because of this, the vast majority of user-initiated trees with a 'false' root node (per 3PFC ratings) go under the radar of 'untrustworthy' classifications (see section S5.3 and figures S14, S27, and S40). User-generated misinformation trees are diffusing a modality of false content different from web domains (e.g., memes, or videos) and, as a result, they are not captured by the 'untrustworthy' label. Because of this, we should consider 'untrustworthy' as an attribute that gives a high recall measure of misinformation, but with low precision. In contrast, 3PFC ratings have perfect precision but unknown recall for misinformation. Given the content moderation policies described in S3.2 and S4, most misinformation does not come from untrustworthy sources. In addition, many large trees from 'untrustworthy' sources likely contain reliable information.

Political Interest Measure We segment users participating in diffusion trees in two groups: users with high political interest, and users with low political interest. High political interest users are those in the top 10% of engagement defined as (a) comments, likes, reactions, or reshares with content classified as political on Facebook during the 90 days prior to our study period (2020-06-01 to 2020-08-31); or (b) views of political content during the same time window. We use these two measurements as alternative operationalizations of the same concept.

The rest of users (bottom 90% in terms of engagement or in terms of views) are considered to have low political interest. We only have this information aggregated at the tree level as the proportion of users that are classified as having high political interest.

Swing State Measure For users participating in each each tree, we also compute the proportion of them that are located in a swing state using their predicted location. Every individual is assigned a residential state based on information and activity on Facebook, including the city reported on Facebook profiles as well as device and connection information (for this, we follow the same approach as prior research, (*8*)). The definition of swing state was based on the two most recent Electoral College Ratings by the Cook Political Report prior to August, 2020. Swing states were those whose complete electoral geography was considered in the "Toss Up", "Lean Democrat", or "Lean Republican" in at least one of the reports. The states that met this criterion are the following: Arizona, Georgia, Maine, and North Carolina ("Toss up" states); and Florida, Michigan, Minnesota, New Hampshire, Pennsylvania, Wisconsin, Iowa, Ohio, Texas ("Lean Democrat" or "Lean Republican" states).

Boosted Content Measure Boosted content is defined as any root post with at least 50% of views attributed to a version of the post that was boosted by its creator[S29]. Approximately 1.5% of the trees in our dataset with 100 or more shares are considered boosted content according to this measure.

S4 Interventions Timeline

Our observation window includes periods of high-intensity interventions designed to moderate content on the platform above and beyond what the Feed algorithm is designed to do [S30]. The leading academic authors reconstructed the timeline of extraordinary platform interventions shown in Figure S11 (including the "break the glass" measures discussed in the main text), using three main sources available in the public record: (1) The memo that circulated among the January 6 House Committee members, leaked to the Washington Post [S31]; (2) BuzzFeed's reporting on the Stop the Steal campaign[S32]; and (3) Jeff Horwitz's reporting in his book *Broken Code. Inside Facebook and the Fight to Expose Its Harmful Secrets*[S33]. The academic team requested, and did not receive from Meta, more details about these platform interventions. According to Meta, they did not provide this information "due to adversarial risks and concerns about the interventions being less effective in the future if their details become public".

[S29] See https://www.facebook.com/business/help/240208966080581?id=352109282177656 for more information about boosted posts, which are ads created from existing posts on a Facebook Page, unlike other types of Facebook ads that are not posts created organically and then boosted.

[S30] https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking.

[S31] https://www.techpolicy.press/read-the-january-6-committee-social-media-report.

[S32] https://www.buzzfeednews.com/article/craigsilverman/facebook-failed-stop-the-steal-insurrection.

[S33] https://www.penguinrandomhouse.com/books/712678/broken-code-by-jeff-horwitz.

Jun, 2020	"An internal Facebook intelligence report warned of growing danger from QAnon activity on Facebook" (J6 Social Media Report, p. 31).
Aug 19, 2020	"Facebook removed thousands of groups, pages, accounts, and ads tied to QAnon and various militia groups and acted to reduce the reach and distribution of remaining accounts and hashtags on the platform" (J6 Social Media Report, p. 32).
Sep 22, 2020	"In September, Nick Clegg had told USA Today that the company had prepared contingency plans but would not be discussing them" (Broken Code, p. 215). - USA Today article: https://www.usatoday.com/story/news/politics/elections/2020/09/22/election-2020-facebook-has-break-glass-measures-if-violence-erupts/5866803002/ - "Facebook' s civic integrity team worked diligently to address competing complex risk areas in advance of the election by developing 63 "break the glass" measures designed to slow the flow of viral, potentially harmful content." (J6 Social Media Report, p. 29).
Sep 30, 2020	By September, the company had stopped "actively boosting non-recommended groups in News Feed" and "established some limits on 'bulk invites'. It also 'imposed a three-week waiting period before newly created groups on any subject were eligible to be recommended to Facebook users" (Broken Code, pp. 209-210). - The 'Groups Task Force' had made a contribution to Civic Integrity's arsenal of Break the Glass interventions: "If all hell broke loose around Election Day, the company could throw a switch and force the administrators of groups with a history of breaking Facebook's rules to manually approve all member posts. The change would both slow down the groups overall and force the admins to be responsible for the content they approved" (Broken Code, p. 210). - During this month, Facebook announced that it would ban political ads before and after the election (Broken Code, p. 211). - During this month, Facebook labeled speech by political figures, directing users toward outside sources of factual information about voting; it then declared the election over when ballots had been counted (Broken Code, p. 211). - During this month, Facebook accepted changing the algorithm so that it would "stop treating anger emojis as grounds to amplify a post" (Broken Code, p. 211).
Oct 4, 2020	Break the glass (BtG) measure to remove all Groups created in the last 21 days from Recommendations in order to offset the low recall and detection of Groups potentially associated with violence and other harms is launched and never deprecated (J6 Social Media Report, p. 36).
Oct 9, 2020	BtG measure called "virality circuit breaker" is deployed to slow the distribution of URLs linking to unknown external domains that may contain misinformation (J6 Social Media Report, p. 35 and p. 57).

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (1 of 8).

Oct 12, 2020	BtG measure to demote videos designated "civic" from news pages with a low News Ecosystem Quality (NEQ) is launched and never deprecated (J6 Social Media Report, p. 35).
Oct 15, 2020	BtG measure limiting to 100 the number of invitations to a group a single user could send, down from an initial limit which may have been as high as 2,250, is launched (J6 Social Media Report, p. 52).
Oct 20, 2020	BtG measures launched: - 'Proportional demotion' for hate speech and graphic violence (J6 Social Media Report, p. 35). - Remove Feed boosts for non-recommendable Groups content (J6 Social Media Report, p. 35 and p. 56). - Remove all civic Groups from recommendations in "Groups you should join" to address low recall of groups associated with real-world harm (J6 Social Media Report, p. 35).
Oct 21, 2021	BtG measure to "freeze commenting on posts in Groups that have a high rate of hate speech and violence and incitement comments" is launched (J6 Social Media Report, p. 35 and p. 57).
Oct 22, 2021	BtG measure to filter low NEQ pages from Pages You May Like in order to prevent misinformation pages from becoming viral is launched (J6 Social Media Report, p. 35).
Oct 26, 2020	BtG measure 'proportional demotion' for violence and incitement is launched (J6 Social Media Report, p. 35).
Nov 3, 2020	Election day.
Nov 4, 2020	The first Stop the Steal Facebook Group is created (J6 Social Media Report, p. 39). The group starts growing quickly: "Super-inviters were doing their thing: 30 percent of the Stop the Steal group's members could be traced to just 0.3 percent of users", Broken Code, p. 219). - "The original Stop the Steal Facebook group was started by pro-Trump activist Kylie Jane Kremer, who said at public events leading up to Jan. 6 that it had more than a million people waiting to be added before it was shut down by the company two days later. The removal of the original group had the effect of cutting off the head of a hydra, as copycat and offshoot groups sprung up in its place." (https://www.buzzfeednews.com/article/craigsilverman/facebook-failed-stop-the-steal-insurrection). - "The original Stop the Steal group, and the offshoots that emerged after it was banned, grew quickly thanks to what the report labels “super-inviter” accounts. The biggest Stop the Steal groups had 137 super-inviters, who invited 67% of the groups’ members, according to the report. These accounts were each responsible for inviting more than 500 people to groups. Facebook’s analysis found the super-inviters worked in coordination, lied about their

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (2 of 8).

	locations, and used private groups and chats to coordinate activity." (https://www.buzzfeednews.com/article/craigsilverman/facebook-failed-stop-the-steal-insurrection).
Nov 5, 2020	On this day, Stop the Steal Facebook Events were scheduled for locations including California, Virginia, Washington DC, Pennsylvania, and Florida (J6 Social Media Report, p. 30). Facebook takes the following measures: - Stop the Steal Group is removed (takedown "splintered into offshoot groups", Broken Code, 233). The group had already accumulated more than 360,000 members and more than 7,000 posts, on which there were more than 200,000 comments (J6 Social Media Report, p. 39). After "the deletion of the first Stop the Steal group, the movement experienced 'meteoric growth, as copycat groups sprung to replace it. At one point, nearly all of the fastest growing civic groups on Facebook were related to Stop the Steal" (J6 Social Media Report, pp. 41-42). - BtG measure to demote posts predicted to be hate speech is launched and kept permanently (J6 Social Media Report, p. 36). - BtG measure to demote content that contains keyword matches for voter fraud or delegitimization claims is launched (J6 Social Media Report, p. 36). - BtG measure to demote content from users who posted multiple pieces of third-party fact-checked misinformation in the past 30 days is launched (J6 Social Media Report, p. 36 and p. 56). - "On the afternoon of November 5, two days after the end of voting, Facebook 'broke glass' on the 2020 U.S. elections. All of a sudden, posts in groups with a history of violating Facebook's rules required admin approval. Users seeking to share election-related content had to click through a notice directing them toward legitimate information sources. Posts classified as having a 70 percent or greater chance of inciting violence began to disappear, and comment threads with abnormally high amounts of hate speech froze. Publishers who scored low on quality metrics received far less distribution, as did posts that were being reshared in long viral chains." (Broken Code, pp. 219-220). - "The title of one Break the Glass measure, "Stop Boosting Content from Non-Recommendable Groups," underscored how remedial all this was. Instead of slowing the platform down in the subtle ways that the company's Integrity staff had long recommended, Facebook was pulling the emergency brake. In total, sixty-four separate break-the-glass measures were in place well before the election was called for Biden on November 7." (Broken Code, p. 220). - "Less than a month later, when Facebook metrics for violence and incitement had simmered down to pre-election levels, the company began rolling them back. Strengthened demotions

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (3 of 8).

of incitement to violence and distribution penalties for users who repeatedly posted misinformation were among the first to go." (Broken Code, p. 220).

- "During the run-up to Election Day, Facebook had removed only lies about the actual voting process—stuff like "Democrats vote on Wednesday" and "People with outstanding parking tickets can't go to the polls." Noting the thin distinction between the claim that votes wouldn't be counted and that they wouldn't be counted accurately, Chakrabarti had pushed to take at least some action against baseless election fraud claims. Civic hadn't won that fight, but with the Stop the Steal group spawning dozens of similarly named copycats—some of which also accrued six-figure memberships—the threat of further organized election delegitimization efforts was obvious." (Broken Code, 233-234).

Nov 7, 2020 Election called for Biden. Facebook's internal security team assessed that the protests were 'losing momentum'" (J6 Social Media Report, pp. 40).

- BtG measure to demote low NEQ news and boost high NEQ news in order to increase the average quality of news in connected news feed is launched (J6 Social Media Report, p. 36 and p. 57).

- "Shortly after the Associated Press called the presidential election for Joe Biden on November 7—the traditional marker for the race being definitively over—Molly Cutler assembled roughly fifteen executives that had been responsible for the company's election preparation. Citing orders from Zuckerberg, she said the election delegitimization monitoring was to immediately stop." (Broken Code, p. 234)

- "(…) the larger threat proved to be to the transition, not the election. In the days after the voting stopped, Facebook saw a significant spike in violence and incitement on the platform (…) As rates [of] misinformation also rose significantly due to false claims of voter fraud, the company rolled out a second suite of 'break glass' measures. These included the use of a 'News Ecosystem Quality (NEQ) score to demote content from untrustworthy news publishers; as much as seventy percent of delegitimizing content from pages came from publishers with low NEQ scores" (J6 Social Media Report, pp. 33-34).

Nov 17, 2020 Zuckerberg testified before the Senate Judiciary Committee (J6 Social Media Report, p. 60).

Nov 30, 2020 "Facebook lifted all demotions of content that delegitimized the election results" (Broken Code, p. 235).

- BtG measure to demote content that contains keyword matches for voter fraud or delegitimization claims is deprecated (J6 Social Media Report, p. 36).

- "Stop the Steal groups proliferated across Facebook between election day and the end of November. Of these, Facebook took action against only 43" (J6 Social Media Report, p. 41).

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (4 of 8).

Dec 1, 2020	"The platform restored misinformation-rich news sources to its "Pages You Might Like" recommendations and lifted a virality circuit breaker" (Broken Code, p. 235). - "Strengthened demotions of incitement to violence and distribution penalties for users who repeatedly posted misinformation were among the first to go." (Broken Code, p. 220). - BtG measure to "freeze commenting on posts in Groups that have a high rate of hate speech and violence and incitement comments" is deprecated (J6 Social Media Report, p. 35 and p. 57).
Dec 2, 2020	The platform "relaxed its suppression of content that promoted violence" (Broken Code, p. 235). Facebook restructures the team behind the break glass measures. On this day, it announces that "the civic integrity team which led Facebook's election preparations would be reorganized into three pillars split across separate teams: one to deal with long-term responses to inauthentic behavior and other harms; one to support this work by creating tools and infrastructures; and one which develops active monitoring and mitigation strategies. This announcement coincided with the departure of several team members from Facebook." (J6 Social Media Report, p. 37).
Dec 3, 2020	The following BtG measures are deprecated: - 'Proportional demotion' for hate speech, violence, and incitement (J6 Social Media Report, p. 35). - Demotion of content from users who posted multiple pieces of third-party fact-checked misinformation in the past 30 days (J6 Social Media Report, p. 36 and p. 56). - Limiting the number of invitations to a group a single user could send to 100 (J6 Social Media Report, p. 52, exact roll-back date unclear in the Report).
Dec 7, 2020	The platform resumed 'Feed boosts for non-recommendable Groups content' (Broken Code, p. 235). The BtG measure to Remove Feed boosts for non-recommendable Groups content is deprecated (J6 Social Media Report, p. 35 and p. 56).
Dec 8, 2020	After this date, at least 34 of the 63 break-glass measures were rolled back. (J6 Social Media Report, p. 34). - "Not all of the break glass measures were rolled back at the same time, and some were not rolled back at all before January 6th. Several measures related to auto-deleting, demoting, and filtering content which might include incitement to violence, for example, were extended multiple times due to concerns about the prevalence of violence and incitement on the platform, especially in comment sections. The rollback process was touch-and-go as staff debated the merits of each measure. There were mistakes: for example, on December 8th one staffer noted that three measures were deactivated 'prematurely due to execution error'; the company declined to reactivate them because they were "not likely to be obvious" and it wasn't worth the risk". These were a freeze on comments in groups with high rates of hateful

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (5 of 8).

	and violent speech; a trigger to auto-disable commenting in group threads with high rates of violent incitement; and a measure to prevent groups from changing their names to delegitimizing terms" (J6 Social Media Report, pp. 34-35).
Dec 10, 2020	The following BtG measures were deprecated: - Filter low NEQ pages from Pages You May Like in order to prevent misinformation pages from becoming viral (J6 Social Media Report, p. 35 and p. 57). - The "virality circuit breaker", deployed to slow the distribution of URLs linking to unknown external domains that may contain misinformation (J6 Social Media Report, p. 36 and p. 57). - Demote low NEQ news and boost high NEQ news in order to increase the average quality of news in connected news feed is launched (J6 Social Media Report, p. 36).
Dec 12, 2020	Violent pro-Trump demonstrations in Washington DC (event in which Proud Boys participated, J6 Social Media Report, p. 38). - https://www.washingtonpost.com/local/trump-dc-rally-maga/2020/12/11/8b5af818-3bdb-11eb-bc68-96af0daae728_story.html
Dec 16, 2020	By this day, Facebook had removed the caps on the bulk group invitations that had driven Stop the Steal's growth (Broken Code, p. 235).
Dec 17, 2020	On this day, "a data scientist flagged that a system responsible for either deleting or restricting high-profile posts that violated Facebook's rules had stopped doing so". (Broken Code, p. 235). - "In fact, the system truly had failed, in early November. Between then and when engineers realized their error in mid-January, the system had given a pass to 3,100 highly viral posts that should have been deleted or labeled 'disturbing'" (Broken Code, p. 235).
Dec 19, 2020	Trump tweets that there would be a "big protest in D.C. on January 6th" (Broken Code, p. 231).
Jan 5, 2021	"By January 5, Facebook was preparing a new crisis coordination team, just in case". (Broken Code, p. 236).
Jan 6, 2021	Facebook starts to restore the safeguards that it had eliminated a month earlier (Broken Code, p. 246). The following BtG measures are relaunched: - 'Proportional demotion' for hate speech, violence, and incitement (J6 Social Media Report, p. 35). - Remove Feed boosts for non-recommendable Groups content (J6 Social Media Report, p. 35 and p. 56).

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (6 of 8).

- Filter low NEQ pages from Pages You May Like in order to prevent misinformation pages from becoming viral (J6 Social Media Report, p. 35 and p. 57).

- Demote content that contains keyword matches for voter fraud or delegitimization claims is relaunched (J6 Social Media Report, p. 36).

- Freeze commenting on posts in Groups that have a high rate of hate speech and violence and incitement comments (J6 Social Media Report, p. 35 and p. 57).

In addition:

- The platform removes two of Trump's posts (Broken Code, p. 237).

- The platform joins YouTube and Twitter in taking down Trump's "We love you. You're very special" video (Broken Code, p. 237).

- Facebook suspends Trump's account for 24 hours.

- "As rioters entered the Senate chamber and offices around the building, while members of Congress donned gas masks and hid where they could, Facebook kept tweaking the platform in ways that might calm things down, going well past the set of Break the Glass interventions that it had rolled out in November. Along with additional measures to slow virality, the company ceased auto-deleting the slur "white trash," which was being used quite a bit as photos of colorfully dressed insurrectionists roaming the Capitol went viral. Facebook had bigger fish to fry than defending rioters from reverse racism" (Broken Code, p. 237).

- "The hashtag "#StopTheSteal was surging in the wake of January 6. No sooner had Integrity teams nuked the hashtag and mapped out networks of advocates using it than they identified a new threat: the same insurrectionist community was uniting to take another shot. The new rallying point was the 'Patriot Party', which pitched itself as a far-right, Trump-supporting alternative to the Republican Party" (Broken Code, p. 237).

Jan 7, 2021 Facebook bars Trump through the end of his term (January 20, date of Biden's inauguration), before deciding to deplatform him indefinitely. (Broken Code, p. 237).

Jan 8, 2021 Facebook delisted "Stop the Steal" from Groups search (in addition to the removal of the term from main search in November, J6 Social Media Report, p. 43).

Jan 11, 2021 Facebook began removing content containing the phrase "Stop the Steal" from the platform (J6 Social Media Report, p. 44).

Jan 13, 2021 BtG measure to demote low NEQ news and boost high NEQ news in order to increase the average quality of news in connected news feed is relaunched (J6 Social Media Report, p. 36).

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (7 of 8).

Jan 14, 2021	BtG measure to demote content from users who posted multiple pieces of third-party fact-checked misinformation in the past 30 days is relaunched (J6 Social Media Report, p. 36 and p. 56).
Jan 22, 2021	"The company's internal state of emergency was finally lowered from its highest level" (Broken Code, p. 240).
Jan 25, 2021	BtG measure 'proportional demotion' for hate speech, violence, and incitement is deprecated (J6 Social Media Report, p. 35).
Jan 29, 2021	The following BtG measures are deprecated: - 'Proportional demotion' for violence and incitement (J6 Social Media Report, p. 35). - Demote of content from users who posted multiple pieces of third-party fact-checked misinformation in the past 30 days (J6 Social Media Report, p. 36 and p. 56). - Freeze commenting on posts in Groups that have a high rate of hate speech and violence and incitement comments (J6 Social Media Report, p. 35 and p. 57).
Jan 30, 2021	BtG measure to demote content that contains keyword matches for voter fraud or delegitimization claims is deprecated (J6 Social Media Report, p. 36).
Feb 16, 2021	The following BtG measures are deprecated: - Filter low NEQ pages from Pages You May Like in order to prevent misinformation pages from becoming viral is deprecated (J6 Social Media Report, p. 36 and p. 57). - Demote low NEQ news and boost high NEQ news in order to increase the average quality of news in connected news feed is launched (J6 Social Media Report, p. 36).
April 5, 2021	BtG measure to remove Feed boosts for non-recommendable Groups content is deprecated (J6 Social Media Report, p. 35 and p. 56).

Figure S11: **Timeline of Interventions.** This timeline was compiled by the leading academic authors using publicly available information (8 of 8).

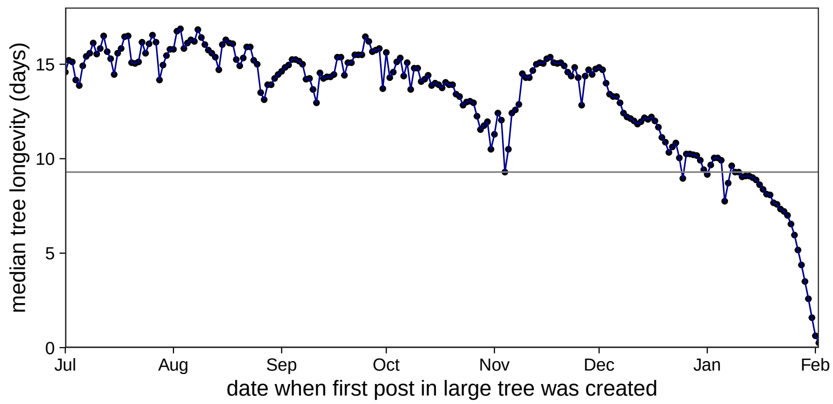

Figure S12: **Tree Longevity for Large Trees.** This figure plots the median tree duration (measured in days) by calendar day during our observation window, i.e., half of the trees initiated on July 1st grew for about two weeks, etc. The horizontal grey line marks the minimum median longevity observed prior to Election Day. The sharp decrease that starts in December is the result of right censoring (i.e., we stopped collecting data as diffusion trees were still growing).

S5 Supplementary Results: Large Trees

In this section, we present additional analyses that complement the results reported in the main text (based on the 1.2% of trees that meet the k => 100 re-shares threshold, accumulating 54.6% of all views). These analyses were registered in the Pre-Analysis Plan (S7). In section S6 we present additional analyses for small trees ($k < 100$ re-shares).

S5.1 Longevity of Large Trees

Figure 1E in the main text shows that $\sim 78\%$ of the trees grow within a week. In Figure S12 we show the median tree longevity (measured in days) by calendar day during our observation period. The horizontal grey line marks the minimum median longevity observed prior to Election Day, which happens on 2020-11-04. The sharp drop in the longevity of trees in the last two weeks results from the censoring that our observation window imposes on the data. As we discuss when presenting figures S31-S41, regression results that account for time and the chronological day of the root post show results consistent to those discussed in the main text. Similarly, we find that the figures we report in the main text remain virtually identical after excluding trees initiated in the last two weeks, which represent approximately 6% of all trees.

S5.2 Changes in Temporal Trends of Large Trees as Percentages

Figure S13 is an extension to Figure 1 in the main paper: instead of counts, here panels G-I track the relative percentage of large trees initiated by Groups, Pages, and Users.

S5.3 Changes in Temporal Trends of Large 'Untrustworthy' Trees

Figure S14 is an extension to Figure 1I in the main paper. Instead of using the 3PFC ratings to identify false content (as described in section S3.2), here we use an alternative way to identify potentially problematic content (described in section S3.4). The 'untrustworthy' label is applied to Pages and Groups that have 2 or more posts rated 'false' by 3PFCs as well as to domains with 2 or more URLs rated 'false' by 3PFCs since the "misinformation repeat offender" program began in 2018. In other words, this label relies on whether the root post of a tree is published by an 'untrustworthy' Page or Group, or whether it contains a URL to an 'untrustworthy' domain. Because of this operationalization, the vast majority of user-initiated trees with a 'false' root post are excluded from the analyses (most user posts do not contain URLs and users themselves do not receive the 'untrustworthy' label in Facebook's systems). This omission is clearly visible in panels A-C of Figure S14, which track tree volume as percentages and counts (panel B shows linear trends fitted to the periods of low and high-intensity interventions, panel C shows 10th degree polynomials fitted before and after election day). Panels D-F plot changes in the views accumulated by 'untrustworthy' trees (with again linear and polynomial models fitted to the data). The vast majority of user-generated trees rated misinformation are not reflected in these data. Users initiated 89% of the $N \sim 114,000$ trees rated 'false', but only $N \sim 16,000$ of these receive also the label 'untrustworthy', with users accounting for only 3% of these trees (which explains why the user category is barely visible in the figure).

Like with misinformation trees, the number of untrustworthy trees systematically declines as the election nears, but the decline is not as dramatic. The counts also rebound towards the end of November but they do not drop after January 6, unlike misinformation trees. Through this comparison, the figure provides additional insight into the impact of content moderation policies on the diffusion dynamics taking place on the platform.

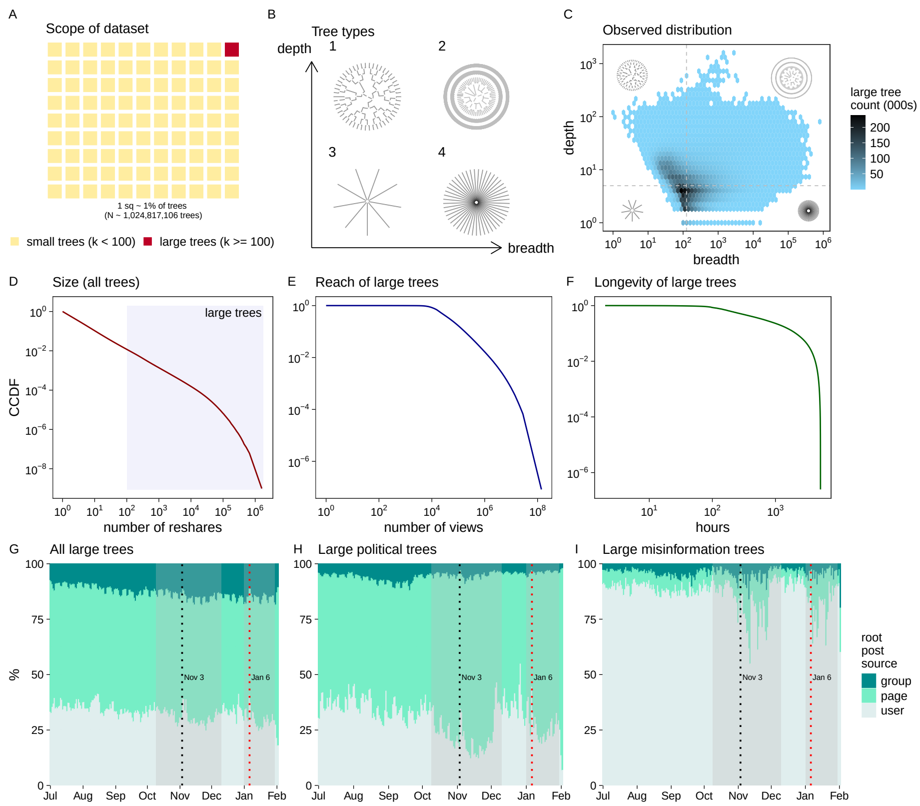

Figure S13: **Description of the Data.** This figure is an extension to Figure 1 in the main text. Panels G-I show changes in percentages, instead of counts. Shaded areas identify high-intensity intervention periods (per the timeline in S11).

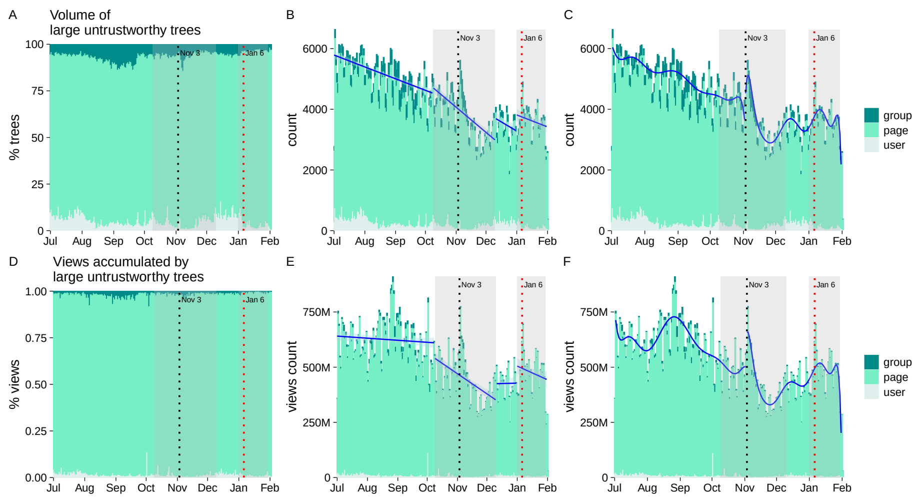

Figure S14: **Changes in the Number of Large Untrustworthy Trees and their Views.** This figure is an extension to Figure 1I in the main text. The time series track changes in the volume of large trees labeled 'untrustworthy' as (A) percentages and (B-C) counts. Blue lines fit four linear models to the data split in the four periods and low and high interventions (in B) and 10th degree polynomials to the data before and after election day (in C). Only trees initiated by posts that contain a URL, or that are published by Pages or in Groups can receive the label 'untrustworthy'. This explains the drastic reduction in this visualization of trees initiated by users, compared to figures 1 and 4 in the main text. The steep drops in late January are an artifact of the right-censoring of the data.

S5.4 Structural Differences across Intervention Periods

Figure S15 complements Figure 2 in the main text by offering the log-log version of the plots. Table S5 displays the summary statistics of the structural properties, and the results of Kolmogorov–Smirnov tests, which confirm that the differences in these distributions are statistically significant (the p-values associated to this test can be interpreted as the estimated likelihood that the null hypothesis of 'no difference' is true).

S5.5 Structural Differences for Large Misinformation Trees

Figure S16 and table S6 complement Figure 2 in the main text by offering summary statistics of the structural properties of the three subsets of trees we examine (misinformation trees, political trees, and all trees) for the full observation period. The upper panels display statistics computed for the full distributions; the lower panels display statistics computed for the size-matched trees. The structural measures we use to characterize the trees (virality, depth, and breadth) are all correlated with tree size, so following prior research (*6*) we run a set of comparisons holding size constant, i.e., comparing only size-matched trees (lower row in Figure S16, see also section S3.1). Even after fixing tree size, trees identified as misinformation are, on average, deeper and have higher virality scores (although there is less fact-checked misinformation among the deeper and most viral trees). In the last two columns of table S6, we also report the results of Kolmogorov–Smirnov tests, which confirm that all differences we report are statistically significant. (As discussed in prior work (*6*) this test is non-parametric and intended for continuous data, yielding conservative estimates for discrete data).

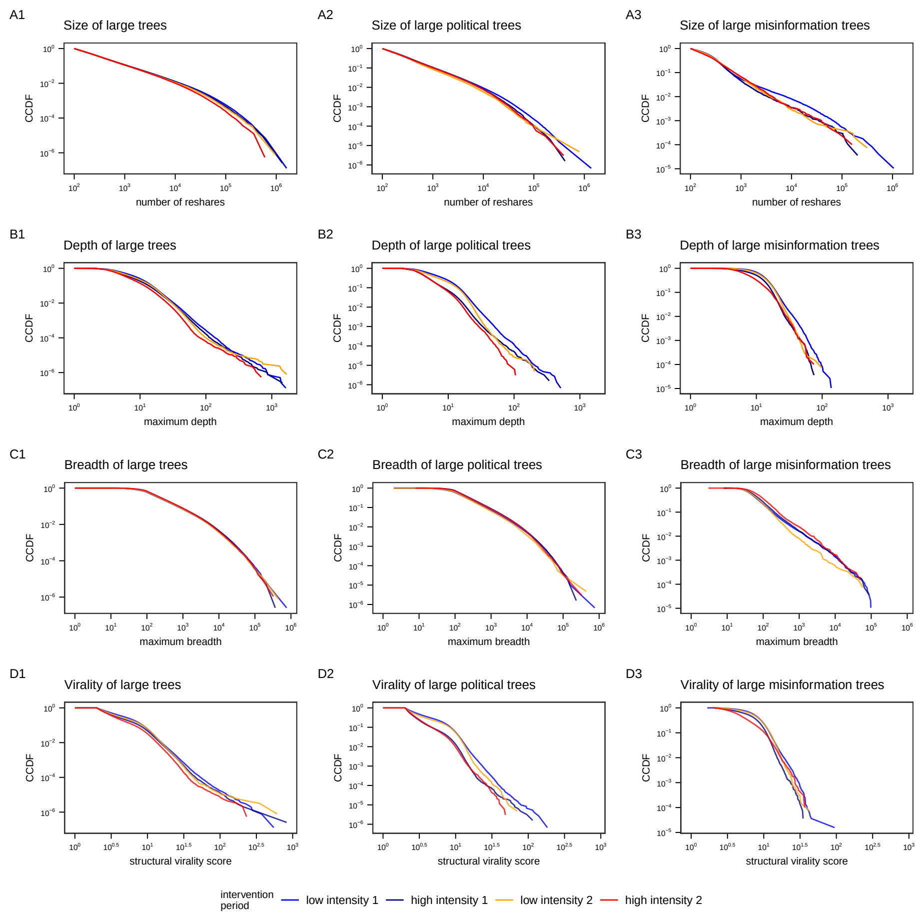

Figure S15: **Structural Differences between All Large Trees, Large Political Trees, and Large Misinformation Trees across Intervention Periods.** This figure is an extension of Figure 2 in the main text. In this version, the horizontal axes are also logged.

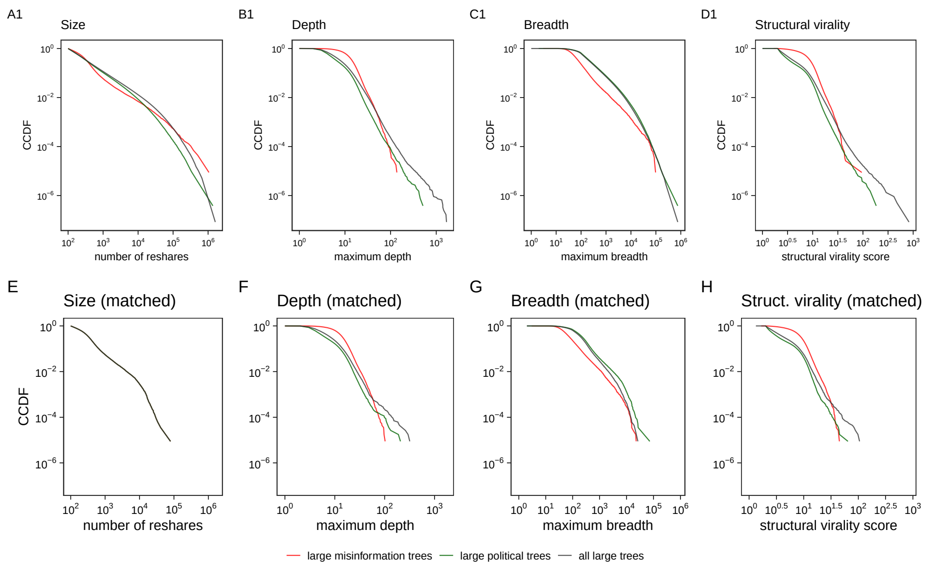

Figure S16: **Structural Differences between Large Misinformation Trees, Large Political Trees, and All Large Trees.** (A-D) These plots summarize the structural properties of three subsets of large trees in our data: trees whose root post was rated false by third-party fact-checkers (misinformation), large trees classified as political, and all large trees. Large trees that were flagged as misinformation are, on average, smaller, so we built a size-matched sample of trees to compare equivalent diffusion events. (E-H) Even after fixing tree size, misinformation trees are, on average, deeper and have higher structural virality, although at each step of the diffusion process, they gather fewer re-shares (hence their lower average breadth). This suggests that misinformation relies less on broadcasting and more on peer-to-peer diffusion through long and narrow paths. See table S6 for additional information, including statistical significance tests.

		Full dataset					
Metric	Subset	p5	p50	p95	Avg	D	p
Size	low intensity 1	105	215	2472	950.9		
	high intensity 1	105	215	2561	928.3	0.01	<1e-6
	low intensity 2	105	217	2444	884.8	0.01	<1e-6
	high intensity 2	105	213	2370	797.7	0.01	<1e-6
Depth	low intensity 1	2	6	19	7.7		
	high intensity 1	2	5	18	6.9	0.09	<1e-6
	low intensity 2	2	5	19	7.1	0.06	<1e-6
	high intensity 2	2	4	16	6.1	0.15	<1e-6
Max Breadth	low intensity 1	27	119	1340	415.2		
	high intensity 1	29	131	1491	442.8	0.05	<1e-6
	low intensity 2	28	128	1402	418.7	0.04	<1e-6
	high intensity 2	33	137	1488	440.8	0.08	<1e-6
Virality	low intensity 1	2.1	3.3	10.5	4.5		
	high intensity 1	2.0	2.9	10.2	4.1	0.08	<1e-6
	low intensity 2	2.0	3.0	10.7	4.3	0.07	<1e-6
	high intensity 2	2.0	2.8	9.1	3.7	0.14	<1e-6

Table S5: **Structural properties of all large trees across intervention periods.** The table displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the distributions to the baseline of the first period of low intensity interventions.

		Full dataset					
Metric	Subset	p5	p50	p95	Avg	D	p
Size	Misinformation	108	247	1145	681.9		
	Political	106	214	2092	697.3	0.08	<1e-6
	All trees	105	215	2477	918.8	0.09	<1e-6
Depth	Misinformation	4	11	23	12.2		
	Political	2	4	16	6	0.55	<1e-6
	All trees	2	5	18	7.2	0.46	<1e-6
Max Breadth	Misinformation	20	53	325	155.6		
	Political	29	134	1535	467.8	0.44	<1e-6
	All trees	28	125	1404	426	0.41	<1e-6
Virality	Misinformation	2.9	7.4	13.6	7.7		
	Political	2.0	2.7	9.3	3.8	0.60	<1e-6
	All trees	2.0	3.1	10.3	4.3	0.52	<1e-6
		Matched subsamples (N=100)					
Metric	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	Misinformation	108	246	1074	439.2		
	Political	108	246	1074	439.2	(<1e-6, <1e-6)	(1, 1)
	All trees	108	246	1074	439.2	(<1e-6, <1e-6)	(1, 1)
Depth	Misinformation	4	11	23	12.1		
	Political	2	4	16	5.9	(0.56, 0.56)	(<1e-6, <1e-6)
	All trees	2	5	18	7.2	(0.44, 0.45)	(<1e-6, <1e-6)
Max Breadth	Misinformation	20	53	304	113.7		
	Political	31	150	808	301.4	(0.47, 0.47)	(<1e-6, <1e-6)
	All trees	29	132	666	239.8	(0.42, 0.42)	(<1e-6, <1e-6)
Virality	Misinformation	2.86	7.36	13.48	7.7		
	Political	2.04	2.65	9.5	3.8	(0.6, 0.6)	(<1e-6, <1e-6)
	All trees	2.05	3.15	10.36	4.3	(0.5, 0.51)	(<1e-6, <1e-6)

Table S6: **Structural properties of large misinformation trees compared to large political trees and all large trees.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.6 Structural Differences by Tree Origin (Large Trees)

Figure S17 expands on Figure 2 in the main text by looking at the structural properties of trees grouped by root node, i.e., whether the tree starts with a user or Page post published in a Group ("group"), a Page post published in its profile ("page"), or a user post in their profile ("user"). The upper row shows the full distributions; the lower row shows the distribution for size-matched trees. Trees that start with a Page post have higher breadth, consistent with broadcasting dynamics (and the larger number of followers that Pages accumulate); but posts published by users are clearly producing deeper and more viral trees. Table S7 confirms that these differences are statistically significant, according to Kolmogorov–Smirnov tests.

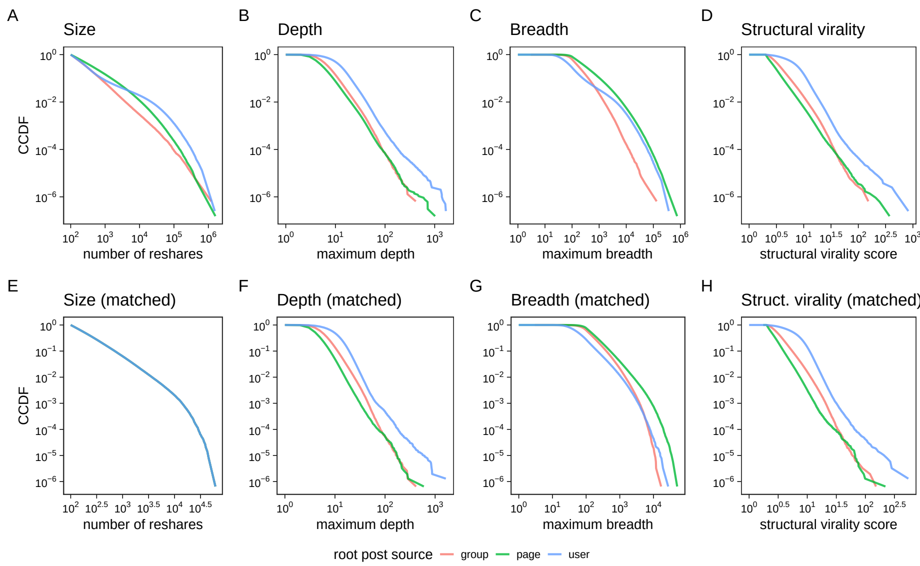

Figure S17: **Structural Differences for Large Trees by Tree Origin.** This figure expands on the information shown in Figure 2 of the main text. The upper row shows the full distributions; the lower row shows the distributions for size-matched trees. The trees with users at the root node grow clearly more viral than the tree initiated by Pages or Groups.

				Full dataset			
Metric	Subset	p5	p50	p95	Avg	D	p
Size	User posts	104	192	2070	1150.1		
	Page posts	107	244	2979	888.7	0.13	<1e-6
	Group posts	104	189	1206	452.8	0.03	<1e-6
Depth	User posts	3	9	25	11.3		
	Page posts	2	4	11	4.9	0.54	<1e-6
	Group posts	2	5	14	6.4	0.40	<1e-6
Max Breadth	User posts	20	59	569	249		
	Page posts	67	183	2000	585.3	0.55	<1e-6
	Group posts	48	120	625	211.6	0.41	<1e-6
Virality	User posts	2.5	6.0	13.3	6.8		
	Page posts	2.0	2.5	5.1	2.9	0.65	<1e-6
	Group posts	2.2	3.1	7.0	3.6	0.51	<1e-6
				Matched subsamples (N=100)			
Metric	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	User posts	104	189	1191	400.6		
	Page posts	104	189	1191	400.6	(<1e-6, <1e-6)	(1, 1)
	Group posts	104	189	1191	400.6	(<1e-6, <1e-6)	(1, 1)
Depth	User posts	3	9	23	10.8		
	Page posts	2	4	9	4.4	(0.58, 0.58)	(<1e-6, <1e-6)
	Group posts	2	5	14	6.3	(0.4, 0.4)	(<1e-6, <1e-6)
Max Breadth	User posts	20	59	366	118.3		
	Page posts	62	145	861	288.4	(0.52, 0.52)	(<1e-6, <1e-6)
	Group posts	48	120	619	205.6	(0.42, 0.42)	(<1e-6, <1e-6)
Virality	User posts	2.48	5.94	12.89	6.6		
	Page posts	2.02	2.45	4.78	2.8	(0.67, 0.67)	(<1e-6, <1e-6)
	Group posts	2.15	3.06	6.99	3.6	(0.5, 0.5)	(<1e-6, <1e-6)

Table S7: **Structural properties of large trees whose root post is a user post compared to large trees whose root is a Page post or a Group post.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.7 Structural Differences by Type of Post (Large Trees)

Boosted posts are organic posts that receive additional distribution (i.e. views) because its original creator paid Facebook to display them as an ad on users' Facebook Feeds. Figures S18 and S19 show the distribution of unmatched and matched trees that have (or have not) a boosted post as the root (about $N \sim 219,000$ trees, or 1.8% of all trees, are initiated by a boosted post). These boosted trees grow less deep and viral than organic posts, and they do have similar broadcasting properties. What these figures suggest, in other words, is that virality does not result, for the most part, from boosted content. The Kolmogorov–Smirnov tests shown in table S8 confirm that these distributions are statistically different. One important caveat here is that the set of posts that are boosted by their creator is likely to be quite different to the set of organic posts that are not boosted, since the process of boosting is driven by selection bias. As a result, this analysis does not allow us to make causal claims about whether boosting a post results, on its own, in larger or deeper trees.

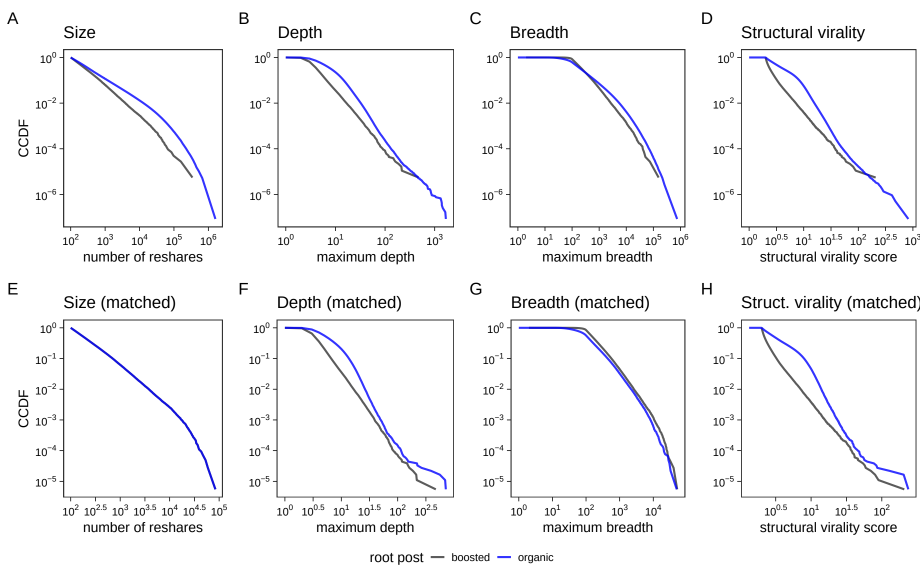

Figure S18: **Structural Differences for Large Trees by Type of Post.** Organic posts generate deeper and more viral trees than boosted posts, and they have similar broadcasting properties.

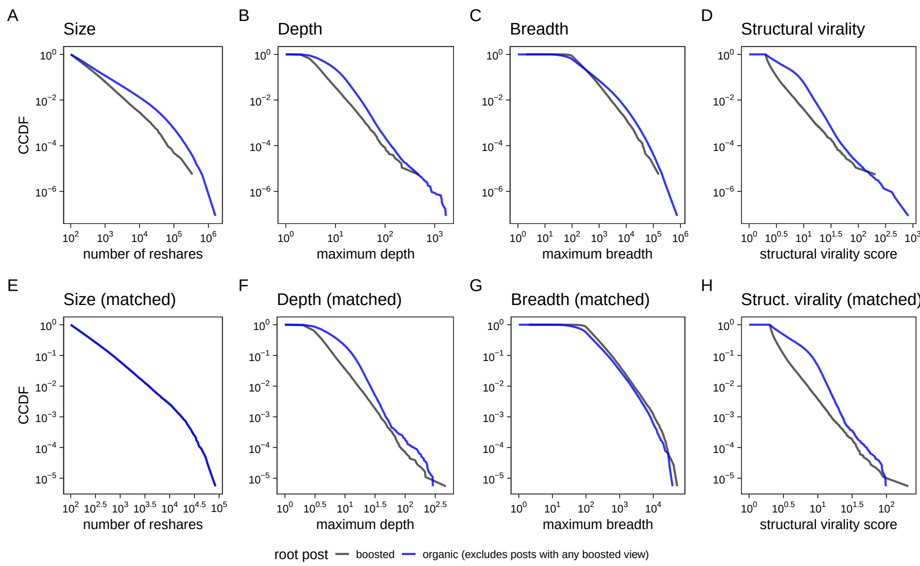

Figure S19: **Structural Differences for Large Trees by Type of Post (Alternative Operationalization).** In order to demonstrate the robustness of the patterns we display in Figure S18, here we exclude trees whose root post received between 0% and 50% boosted views. Organic posts generate deeper and more viral trees than boosted posts also when trees with any boosted views are excluded from the counts of organic posts.

		Full dataset					
Metric	Subset	p5	p50	p95	Avg	D	p
Size	Organic root post	105	216	2503	926.3		
	Boosted root post	104	184	1188	430	0.09	<1e-6
Depth	Organic root post	2	5	18	7.3		
	Boosted root post	2	3	8	3.8	0.38	<1e-6
Max Breadth	Organic root post	28	125	1413	427.5		
	Boosted root post	76	156	934	325.9	0.24	<1e-6
Virality	Organic root post	2.1	3.1	10.3	4.3		
	Boosted root post	2.0	2.2	4.0	2.5	0.42	<1e-6
		Matched subsamples (N=100)					
Metric	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	Organic root post	104	184	1183	409.5		
	Boosted root post	104	184	1183	409.5	(<1e-6, <1e-6)	(1, 1)
Depth	Organic root post	2	5	17	6.7		
	Boosted root post	2	3	8	3.8	(0.35, 0.36)	(<1e-6, <1e-6)
Max Breadth	Organic root post	26	110	745	233.5		
	Boosted root post	76	156	930	316.9	(0.29, 0.29)	(<1e-6, <1e-6)
Virality	Organic root post	2.05	3.05	9.78	4.2		
	Boosted root post	2	2.19	4.03	2.5	(0.41, 0.41)	(<1e-6, <1e-6)

Table S8: **Structural properties of large trees with organic root posts compared to large trees with boosted root posts.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.8 Structural Differences by 3PFC Ratings (Large Trees)

We also compared the structural properties of trees across the three categories used in the fact-checking program (see S3.2 for a list of fact-checking organizations involved in this program). As mentioned, most of the trees in our data are unrated ($N \sim 12M$). Only $N \sim 114,000$ trees (0.9%) are labelled as misinformation, $N \sim 89,000$ are labelled as misleading, and $N \sim 2200$ are labeled as non-misinformation. In other words, most of the posts with 100 or shares or more evaluated by fact-checkers are corroborated as false or deceptive. In Figure S20 we show that there are no visible structural differences when trees are matched by size (lower row), although misinformation and misleading trees grow larger (upper row). This is not entirely surprising given that this is the small subset of trees selected for inspection by 3PFCs, and posts that receive more diffusion (i.e., posts that trigger larger trees) are more likely to be inspected. In table S9 we show the results of Kolmogorov–Smirnov tests comparing the distribution of trees labeled false with the distribution of misleading trees, and trees rated true.

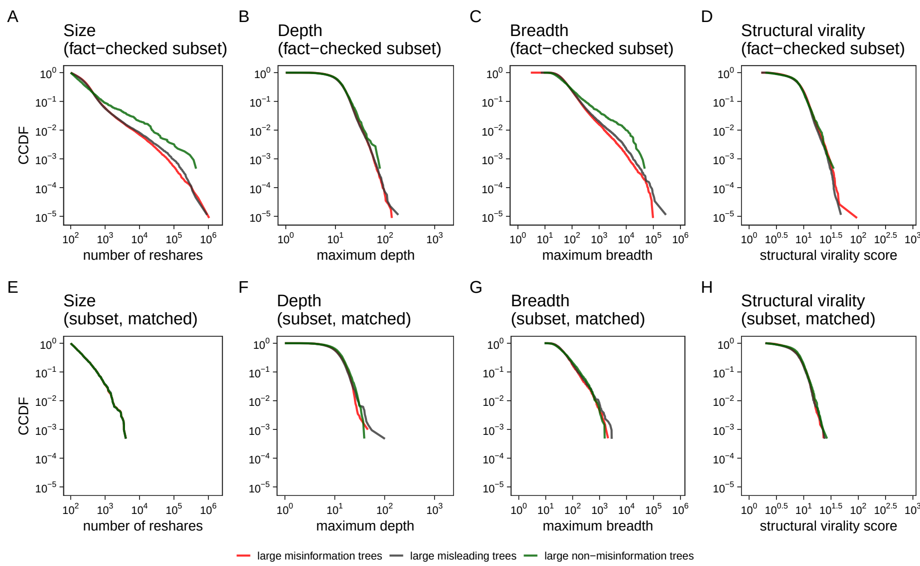

Figure S20: **Structural Differences for Large Trees by 3PFC Ratings.** These plots are based on the subset of trees that were evaluated by third-party fact-checkers. Misinformation and misleading trees grow larger and deeper than trees rated as containing no misinformation (upper row). However, there are no visible structural differences when trees are matched by size (lower row).

Metric	Subset	p5	p50	p95	Avg	D	p
			Full dataset				
Size	3PFC rating: false	108	247	1145	681.9		
	3PFC rating: misleading	108	238	1135	768	0.02	<0.01
	3PFC rating: true	104	187	2688.45	1626.2	0.15	<0.01
Depth	3PFC rating: false	4	11	23	12.2		
	3PFC rating: misleading	4	11	22	12	0.02	0.00
	3PFC rating: true	4	11	25	12.5	0.02	0.41
Max Breadth	3PFC rating: false	20	53	325	155.6		
	3PFC rating: misleading	20	52	363	194.8	0.01	0.00
	3PFC rating: true	17	47	823.9	390.5	0.11	<0.01
Virality	3PFC rating: false	2.9	7.4	13.6	7.7		
	3PFC rating: misleading	2.7	7.3	13.2	7.5	0.02	<0.01
	3PFC rating: true	2.5	7.5	13.6	7.6	0.04	<0.01
			Matched subsamples (N=100)				
Metric	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	3PFC rating: false	104	179	848.6	291.2		
	3PFC rating: misleading	104	179	849.9	291.2	(0.0037, 0.011)	(1, 1)
	3PFC rating: true	104	179	849	291.2	(0.0019, 0.0055)	(1, 1)
Depth	3PFC rating: false	4	10	21	11.2		
	3PFC rating: misleading	4	10	20	11.1	(0.013, 0.037)	(0.12, 1)
	3PFC rating: true	4	11	22	12	(0.067, 0.099)	(<1e-6, 0.00018)
Max Breadth	3PFC rating: false	18	45	235.6	80.7		
	3PFC rating: misleading	18	45	278.4	87.8	(0.015, 0.035)	(0.16, 0.98)
	3PFC rating: true	17	44	309	87.4	(0.047, 0.069)	(0.00011, 0.023)
Virality	3PFC rating: false	2.86	6.95	12.66	7.2		
	3PFC rating: misleading	2.71	6.85	12.4	7	(0.018, 0.041)	(0.058, 0.9)
	3PFC rating: true	2.63	7.49	12.84	7.5	(0.074, 0.11)	(<1e-6, 2.8e-05)

Table S9: **Structural properties of large trees rated as false compared to large trees rated as misleading and large trees rated as true.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.9 Structural Differences for COVID-related Posts (Large Trees)

In Figure S21 we compare COVID-related trees that are classified as misinformation with all other COVID-trees (for reference, we also add the distributions for all trees minus those classified in the first two groups). The results echo what we discuss in Figure 2 of the main text: misinformation trees do not grow as large as non-misinformation trees, but they grow deeper and have higher virality. The lower breath suggests they do not spread via broadcasting but through narrow paths of users re-sharing users. In table S10 we display the results of Kolmogorov–Smirnov tests showing that the differences between these distributions are statistically significant.

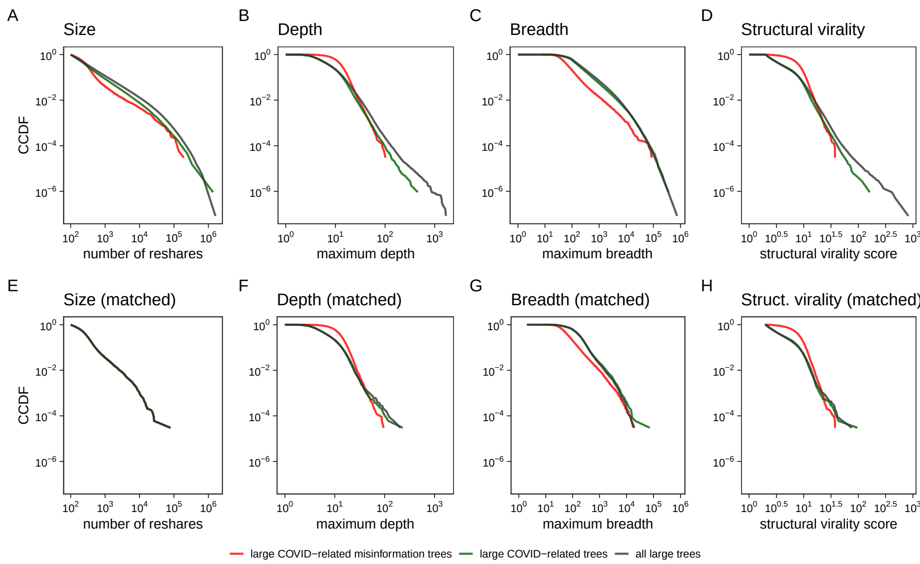

Figure S21: **Structural Differences for Large COVID Trees.** These plots are based on the subset of trees that were classified as COVD-related and, within this category, those labeled as 'misinformation' vs the rest (for reference, we also show the distribution for all trees). These patterns echo what we show in Figure 2 in the main text: misinformation trees grow less and they do not rely on broadcasting but on narrow, deep paths of users re-sharing users.

		Full dataset					
Metric	Subset	p5	p50	p95	Avg	D	p
Size	COVID misinformation	108	227	827	476		
	COVID-related	105	201	1686	650.8	0.09	<1e-6
	All trees	105	215	2477	918.8	0.11	<1e-6
Depth	COVID misinformation	4	11	21	11.4		
	COVID-related	2	5	17	6.9	0.46	<1e-6
	All trees	2	5	18	7.2	0.46	<1e-6
Max Breadth	COVID misinformation	21	50	288	136.2		
	COVID-related	26	114	1081	369.8	0.40	<1e-6
	All trees	28	125	1404	426	0.44	<1e-6
Virality	COVID misinformation	2.9	7.0	12.3	7.2		
	COVID-related	2.1	3.0	9.9	4.2	0.51	<1e-6
	All trees	2.0	3.1	10.3	4.3	0.52	<1e-6
		Matched subsamples (N=100)					
Metric	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	COVID misinformation	108	226	779	348.6		
	COVID-related	108	226	779	348.6	(<1e-6, <1e-6)	(1, 1)
	All trees	108	226	779	348.6	(<1e-6, <1e-6)	(1, 1)
Depth	COVID misinformation	4	11	20	11.3		
	COVID-related	2	5	16	6.7	(0.47, 0.48)	(<1e-6, <1e-6)
	All trees	2	5	17	6.7	(0.47, 0.48)	(<1e-6, <1e-6)
Max Breadth	COVID misinformation	21	50	269	101.8		
	COVID-related	28	123	535	211.7	(0.42, 0.43)	(<1e-6, <1e-6)
	All trees	29	126	523	204.4	(0.44, 0.45)	(<1e-6, <1e-6)
Virality	COVID misinformation	2.9	7	12.2	7.2		
	COVID-related	2.06	3.01	9.89	4.2	(0.51, 0.52)	(<1e-6, <1e-6)
	All trees	2.05	3.03	10	4.2	(0.52, 0.53)	(<1e-6, <1e-6)

Table S10: **Structural properties of large COVID-related misinformation trees compared to large COVID-related trees and all large trees.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.10 Structural Differences by Users Ideology (Large Trees)

In figure S22 we compare political and non-political trees, split by whether a majority of the users that are part of the tree (i.e. they re-shared the root post) are predicted to be liberal or predicted to be conservative. We find that liberal-majority trees (political and non-political) are larger en size, deeper, and more viral than conservative-majority trees. However, political conservative-majority trees tend to be broader. In table S11 we display the results of Kolmogorov–Smirnov tests showing that the differences between these distributions are statistically significant.

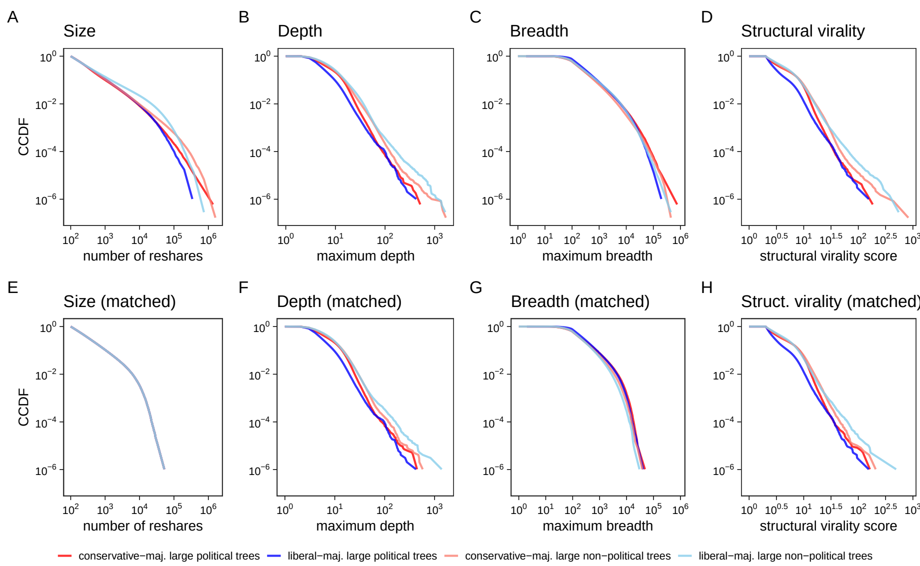

Figure S22: **Structural Differences for Large Trees by Ideology of Users in Tree.** These plots compare the subsets of trees where conservative vs liberal users account for a majority of the re-shares within the tree. We also disaggregate the trees by whether they are classified as political or not.

					Full dataset			
Metric	Ideology	Subset	p5	p50	p95	Avg	D	p
Size	Conservative	Political	106	215	2035	694		
	Liberal	Political	105	214	2186	703.7	0.02	<1e-6
	Conservative	Not political	105	215	2186	834.9	0.02	<1e-6
	Liberal	Not political	105	215	3570	1227.4	0.04	<1e-6
Depth	Conservative	Political	2	5	17	6.6		
	Liberal	Political	2	4	12	5	0.15	<1e-6
	Conservative	Not political	2	5	19	7.3	0.08	<1e-6
	Liberal	Not political	2	6	20	7.9	0.11	<1e-6
Max Breadth	Conservative	Political	26	121	1510	459		
	Liberal	Political	40	156	1580	483.7	0.15	<1e-6
	Conservative	Not political	26	122	1243	389.2	0.02	<1e-6
	Liberal	Not political	29	125	1609	457.2	0.02	<1e-6
Virality	Conservative	Political	2.0	2.9	10.1	4.1		
	Liberal	Political	2.0	2.5	7.0	3.1	0.17	<1e-6
	Conservative	Not political	2.1	3.1	10.7	4.4	0.08	<1e-6
	Liberal	Not political	2.1	3.3	10.3	4.5	0.10	<1e-6
					Matched subsamples (N=100)			
Metric	Ideology	Subset	p5	p50	p95	Avg	D (p5, p95)	p (p5, p95)
Size	Conservative	Political	105	212	1964	544.6		
	Liberal	Political	105	212	1964	544.6	(<1e-6, <1e-6)	(1, 1)
	Conservative	Not political	105	212	1964	544.6	(<1e-6, <1e-6)	(1, 1)
	Liberal	Not political	105	212	1964	544.6	(<1e-6, <1e-6)	(1, 1)
Depth	Conservative	Political	2	5	17	6.6		
	Liberal	Political	2	4	12	4.9	(0.15, 0.15)	(<1e-6, <1e-6)
	Conservative	Not political	2	5	18	7.2	(0.077, 0.08)	(<1e-6, <1e-6)
	Liberal	Not political	2	6	18	7.4	(0.11, 0.11)	(<1e-6, <1e-6)
Max Breadth	Conservative	Political	26	121	1462	372		
	Liberal	Political	40	155	1428	391.1	(0.15, 0.15)	(<1e-6, <1e-6)
	Conservative	Not political	26	121	1131	308.9	(0.024, 0.025)	(<1e-6, <1e-6)
	Liberal	Not political	29	122	1021	281.4	(0.03, 0.03)	(<1e-6, <1e-6)
Virality	Conservative	Political	2.04	2.86	10.06	4.1		
	Liberal	Political	2.04	2.46	6.97	3.1	(0.17, 0.17)	(<1e-6, <1e-6)
	Conservative	Not political	2.05	3.1	10.57	4.3	(0.082, 0.084)	(<1e-6, <1e-6)
	Liberal	Not political	2.05	3.27	9.73	4.4	(0.1, 0.1)	(<1e-6, <1e-6)

Table S11: **Structural properties of large trees with a majority of shares created by conservative users compared with large trees with a majority of shares by liberal users, within subsets of large trees classified as political and not political.** The first panel displays summary statistics (5th, 50th, and 95th percentiles and average) for four key tree-level metrics, as well as the D statistic and p-value for a KS test comparing the full, unmatched sample. The second panel displays the same metrics for a single (randomly selected) matched sample, as well as the 5th and 95th percentiles of the distribution of D statistics and p-values across KS tests performed on each of the 100 matched samples.

S5.11 Speed of Growth (Large Trees)

Figure S23 plots the time it takes for the trees to grow measured in days (main plots) and in hours (insets, first 24 h). We divide trees in three groups: trees classified as misinformation; trees classified as political (excluding misinformation); and the rest of the trees (excluding misinformation and political content). The figure shows that misinformation trees grow more slowly, but they attain greater depth faster, at about double the speed. Since users account for the large majority of re-shares in the trees we analyze, this suggests that misinformation travels further from the original poster in the underlying social network (i.e., it relies more on friends of friends of friends etc to achieve the same diffusion size as non-misinformation). The same dynamics appear with COVID-related content, as shown in Figure S24: misinformation trees grow more slowly, but they reach greater depth much faster.

S5.12 Structural Properties over Time (Large Trees)

In figures S25 and S26 we show changes in the structural properties of trees as a function of the date when the root post was published, the type of content in the post (political, misinformation), and whether the source is a Page, a Group or a user post. Most of the findings reported in the main text (especially Figure 2) hold longitudinally: misinformation trees do not grow as large as political trees (or all other trees), but they grow deeper and more viral (with dips on these two statistics around election day and January 6). Note that these dips are also visible for political content posted by users.

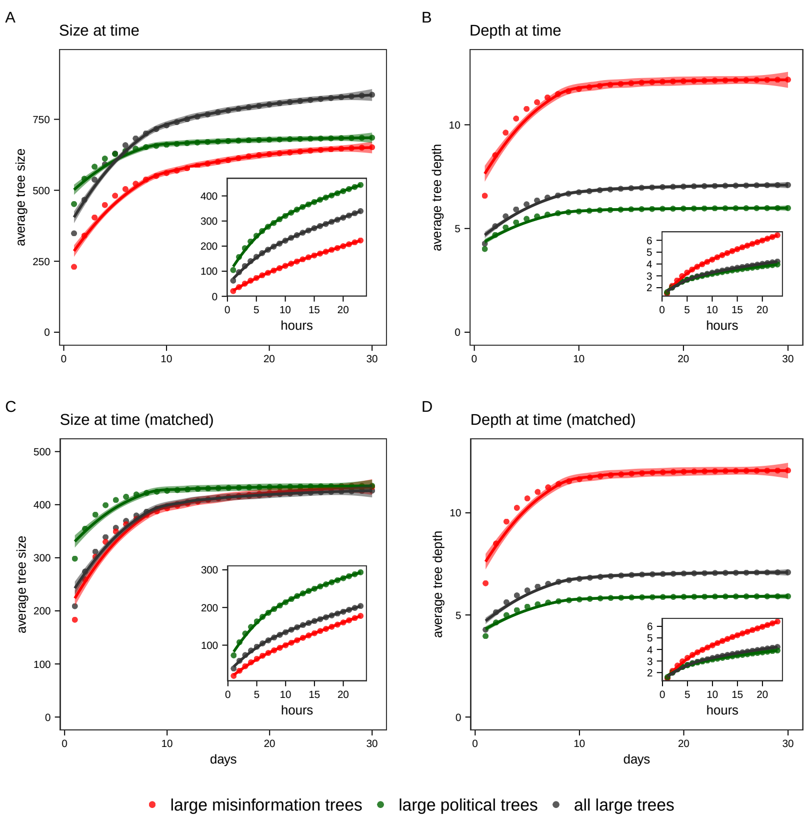

Figure S23: **Speed of Growth for Large Trees.** These plots compare the time it takes for trees to grow. The upper row plots are based on all the trees; the lower row plots rely on the size-matched dataset. Misinformation trees grow more slowly but they attain higher depth faster (at about double speed). This pattern indicates that this type of content travels further away from the original poster in the underlying social network through narrow transmission paths.

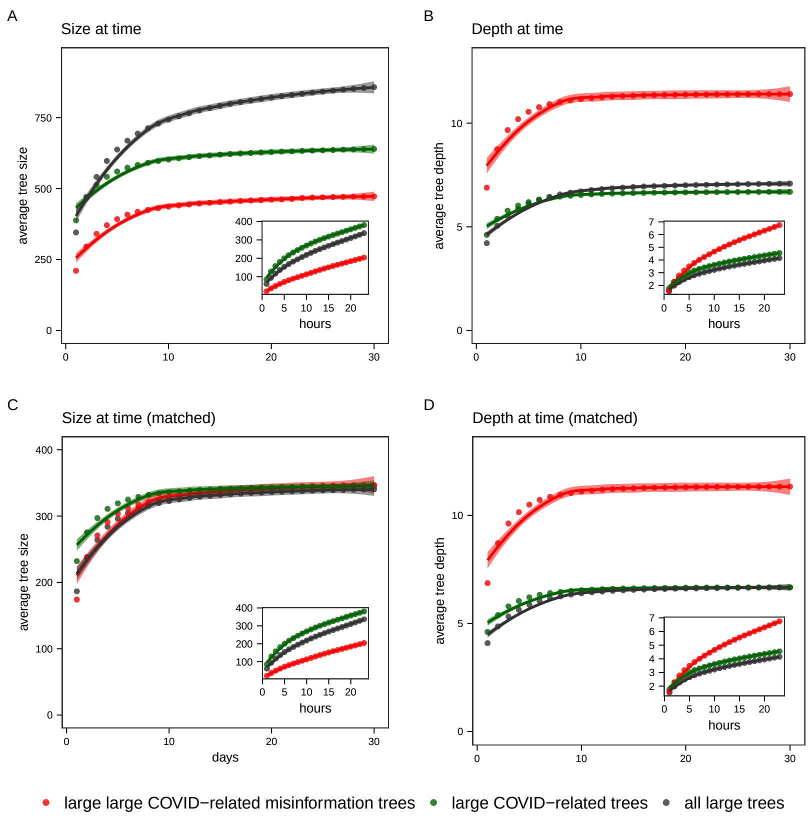

Figure S24: **Speed of Growth for Large COVID-related Trees.** With COVID-related content, misinformation trees also grow more slowly and also attain higher depth faster.

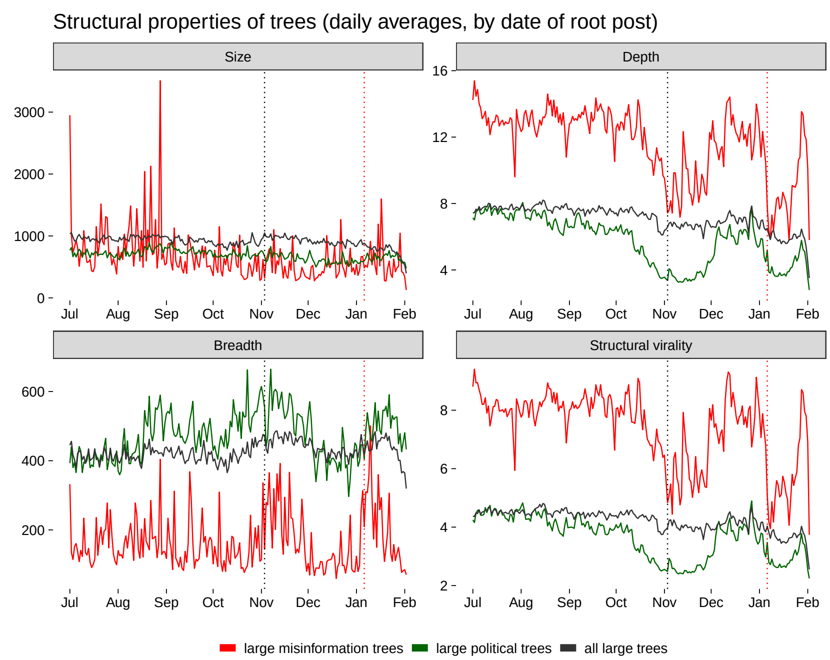

Figure S25: **Structural Properties of Large Trees Depending on Publication Day of Root Post.** This figure is an extension to Figure 2 in the main text. Misinformation trees are deeper and more viral during most of the observation period, with clear dips around election day and after January 6.

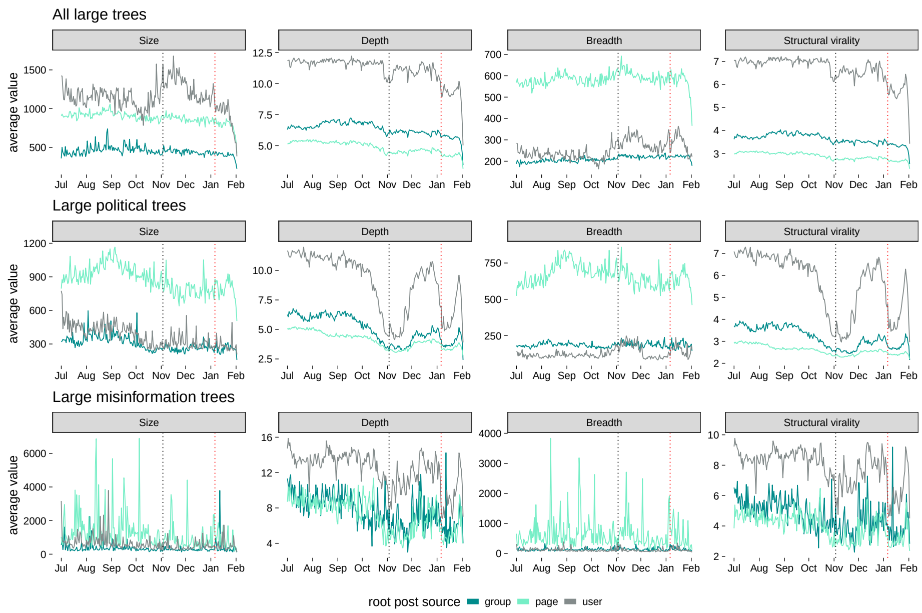

Figure S26: **Structural Properties of Large Trees Depending on Publication Day and Source of Root Post.** This Figure is an extension to Figure 2 in the main text and S25. Misinformation trees are deeper and more viral during most of the observation window, with clear dips around the election date and after January 6. The depth and virality of political trees and misinformation trees decreased significantly around election day and after January 6.

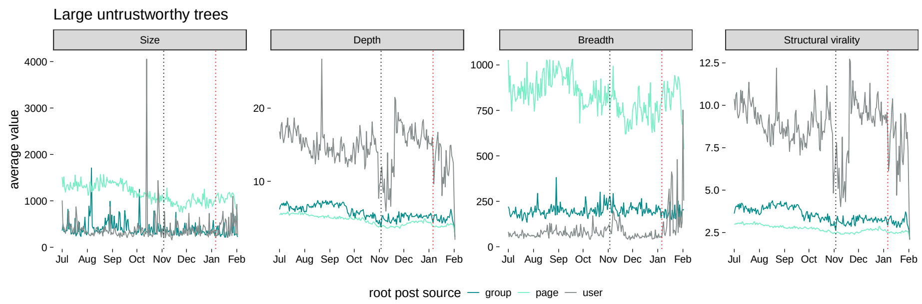

Figure S27: **Structural Properties of Large Untrustworthy Trees Depending on Publication Day and Source of Root Post.** This figure is an extension to figures S25 and S26. Like misinformation, untrustworthy trees initiated by users are deeper and more viral during most of the observation period, with a clear dip around election day (no visible dip after January 6). See also section S5.3 and Figure S14 for more context on the coverage of the 'untrustworthy' label.

S5.13 User Composition (Large Trees)

In addition to the structural and dynamic properties summarized in the previous three subsections, we also have user-level data (aggregated at the tree-level) that allows us to differentiate trees in terms of their composition. In particular, we have information about the ideology, age, level of political interest, and location of the users participating in each diffusion tree, as well the fraction of shares in each tree generated by Pages, Groups, and users.

Figure S28 plots the distribution of each of these variables for all trees, for political trees, and for trees classified as misinformation. Panel A shows the distribution of trees according to their ideological composition, i.e, the ratio $(C-L)/(C+L)$, where C stands for conservative and L stands for liberal (see S3.3 for details on the ideology classifier). While the ideological score distribution for all trees is, overall, quite uniform, political trees are clearly polarized: the diffusion of political content takes place mostly among liberal users or among conservative users, but there are very few posts that get re-shares from the two groups; in addition, Figure S28, panel A shows that misinformation trees diffuse predominantly among conservative users, revealing the same asymmetry identified in prior work (*5*). Panel B shows the average age of users participating in the trees: older users are behind most diffusion of political content and misinformation. Panels C and D summarize tree composition in terms of levels of political interest measured as engagement with and views of political content (other than generating trees with size $k > 100$ shares, see S3.3). Unsurprisingly, political trees and misinformation trees tend to grow disproportionately more via re-shares from users with high levels of political interest. Panel E suggests that composition in terms of location does not change substantially across tree categories, although there is a higher number of misinformation trees with a higher fraction of users in swing states. Panel F shows that Pages are rarely responsible for shares in the trees (see also Figure S17 for cases where Pages are the root node). In table S12 we display the results of non-parametric permutation tests showing that the differences between these distributions are statistically significant. Similar distributions appear with COVID-related content, as shown in Figure S29 and table S13.

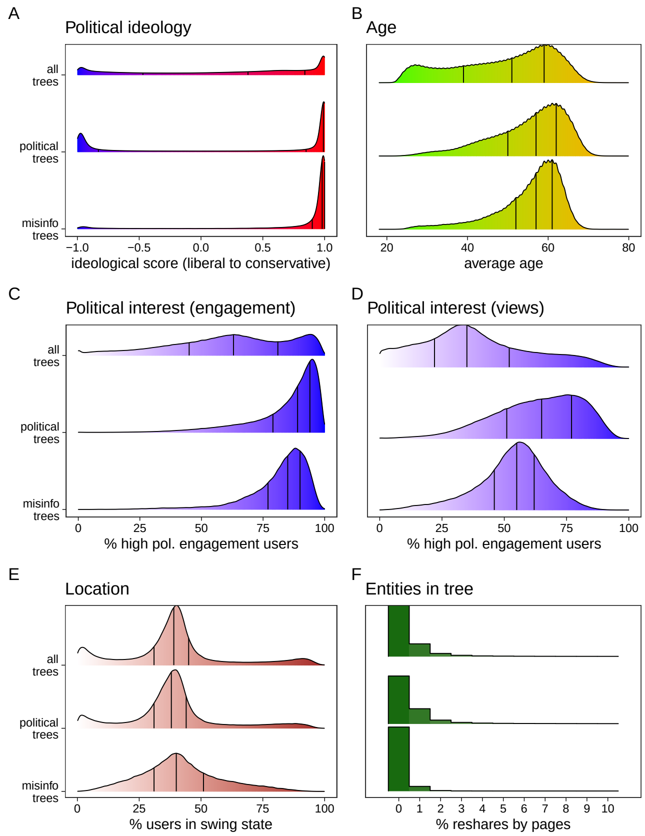

Figure S28: **Tree Composition for Political Posts (Large Trees).** (A) The ideological score of trees, measured as the ratio $(C - L)/(C + L)$, where C stands for conservative and L stands for liberal. The distribution of this score for all trees is, overall, quite uniform, but political trees are clearly polarized, revealing an asymmetry that is much starker for misinformation trees, which diffuse predominantly among conservative users. (B) Older users are behind most diffusion of political content and misinformation. (C-D) Political and misinformation trees are created by users with higher levels of political interest. (E) Tree composition in terms of location does not change substantially across tree categories, although a higher number of misinformation trees have a higher fraction of users in swing states. (F) Pages are rarely responsible for shares in the trees (see Figure S17 for cases where Pages are the root node of the trees).

Ideological score (liberal to conservative)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	-0.64	0.98	1	0.79			
Political	-0.99	0.85	1	0.26	-0.53	(-0.54, -0.53)	<1e-6
All trees	-0.96	0.38	1	0.19	-0.6	(-0.60, -0.60)	<1e-6
Average age of users in tree							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	38	57	65	55.27			
Political	37	57	67	55.16	-0.11	(-0.16, -0.05)	<1e-6
All trees	27	51	65	48.77	-6.5	(-6.54, -6.45)	<1e-6
% high political interest users (engagement)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	56	85	95	81.36			
Political	53	89	98	84.11	2.75	(2.68, 2.83)	<1e-6
All trees	15	63	95	61.02	-20.34	(-20.40, -20.26)	<1e-6
% high political interest users (views)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	26	55	74	53.08			
Political	32	65	87	63.12	10.04	(9.94, 10.13)	<1e-6
All trees	5	35	79	38.02	-15.06	(-15.13, -14.99)	<1e-6
% users in swing state							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	16	40	73	41.8			
Political	5	38	81	38.9	-2.9	(-2.99, -2.78)	<1e-6
All trees	3	39	87	39.13	-2.67	(-2.77, -2.59)	<1e-6
% reshares by Pages							
Subset	p5	p50	p95	Avg	diff	95% CI	p
Misinformation	0	0	1	0.09			
Political	0	0	2	0.54	0.45	(0.46, 0.46)	<1e-6
All trees	0	0	2	0.46	0.37	(0.37, 0.38)	<1e-6

Table S12: **Demographic composition of large misinformation trees compared to large political trees and all large trees.** This table displays summary statistics (5th, 50th, and 95th percentiles and average) for the distribution of demographic characteristics of the users who created the posts in each subset of trees, as well as whether their average values are different for misinformation trees compared to political trees and to all trees. The last two columns report 95% non-parametric bootstrapped confidence intervals for the difference and the p-value resulting from a non-parametric permutation test.

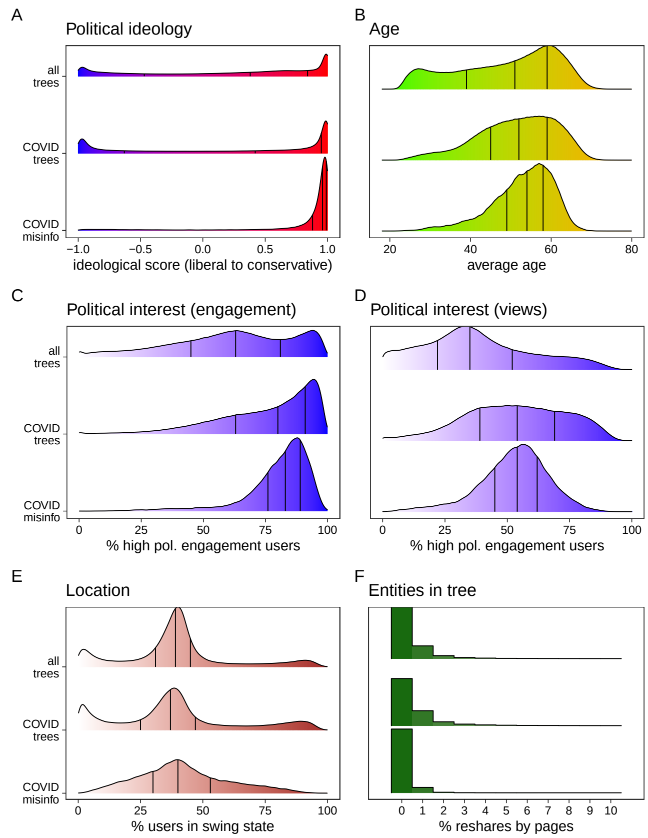

Figure S29: **Tree Composition for COVID Posts (Large Trees).** (A) The ideological score of trees, measured as the ratio $(C - L)/(C + L)$, where C stands for conservative and L stands for liberal. The distribution of this score for all trees is, overall, quite uniform, but COVID-related trees are clearly polarized, revealing an asymmetry that is, again, much starker for misinformation trees. (B) Older users are behind most diffusion of COVID-related content and misinformation (although slightly younger than for political posts). (C-D) Again, COVID-related trees are generated by users with higher levels of political interest. (E) Tree composition in terms of location does not change substantially across tree categories. (F) Pages are rarely responsible for shares in the trees (see Figure S17 for cases where Pages are the root node of the trees).

Ideological score (liberal to conservative)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	-0.33	0.96	1	0.81			
COVID-related	-0.98	0.42	1	0.19	-0.62	(-0.63, -0.62)	<1e-6
All trees	-0.96	0.37	1	0.19	-0.62	(-0.63, -0.62)	<1e-6
Average age of users in tree							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	38	54	63	52.91			
COVID-related	34	52	65	51.43	-1.48	(-1.56, -1.40)	<1e-6
All trees	27	51	65	48.72	-4.19	(-4.26, -4.11)	<1e-6
% high political interest users (engagement)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	52	83	94	80.15			
COVID-related	38	80	97	75.17	-4.98	(-5.11, -4.85)	<1e-6
All trees	15	63	95	60.88	-19.27	(-19.43, -19.15)	<1e-6
% high political interest users (views)							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	27	54	74	52.69			
COVID-related	19	54	84	53.33	0.64	(0.52, 0.77)	<1e-6
All trees	5	35	79	37.91	-14.78	(-14.93, -14.63)	<1e-6
% users in swing state							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	14	40	76	42.12			
COVID-related	2	37	88	38.76	-3.36	(-3.56, -3.20)	<1e-6
All trees	3	39	87	39.11	-3.01	(-3.20, -2.80)	<1e-6
% reshares by Pages							
Subset	p5	p50	p95	Avg	diff	95% CI	p
COVID misinformation	0	0	1	0.1			
COVID-related	0	0	3	0.64	0.54	(0.53, 0.54)	<1e-6
All trees	0	0	2	0.46	0.36	(0.35, 0.36)	<1e-6

Table S13: **Demographic composition of large COVID-related misinformation trees compared to large COVID-related trees and all large trees.** This table displays summary statistics (5th, 50th, and 95th percentiles and average) for the distribution of demographic characteristics of the users who created the posts in each subset of trees, as well as whether their average values are different for COVID-related misinformation trees compared to COVID-related trees and to all trees. The last two columns report 95% non-parametric bootstrapped confidence intervals for the difference and the p-value resulting from a non-parametric permutation test.

S5.14 Regression Models (Large Trees)

In this section we present alternative specifications to the regression models discussed in Figure 3 of the main text. In particular, we run the same regression for the full time period (S30), subsets of data based on content (i.e., political trees: S33 and S35; misinformation trees: S37 and S39; and untrustworthy trees: S40); distributional properties (quantile regression: S31, S34, S38); and different time periods (before/after election; before after January 6: S32, S36, S41). Tables S14, S15, and S16 show the OLS regression estimates with date fixed effects. In general, these alternative specifications are consistent with the results discussed in the main text.

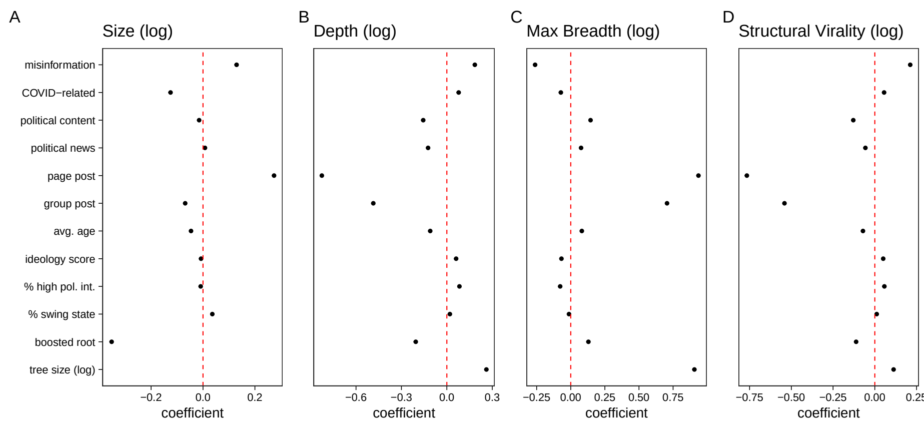

Figure S30: **Correlates of Large Diffusion Trees** These panels show the results of OLS regressions with four dependent variables (tree size, depth, maximum breadth, and structural virality) and for three partitions of the data (large misinformation trees, large political trees, and all large trees). To aid interpretability, all continuous variables have been standardized (with the exception of tree size, which is logged). The models also include daily fixed effects, to control for underlying but unobservable factors that may be changing in time. The trees that grow larger are posted by Pages (which have a larger audience and higher broadcasting potential, also indicated by the higher breadth of their trees); but, controlling by whether the root is a Page post, misinformation trees are clearly larger (A) and more viral (D).

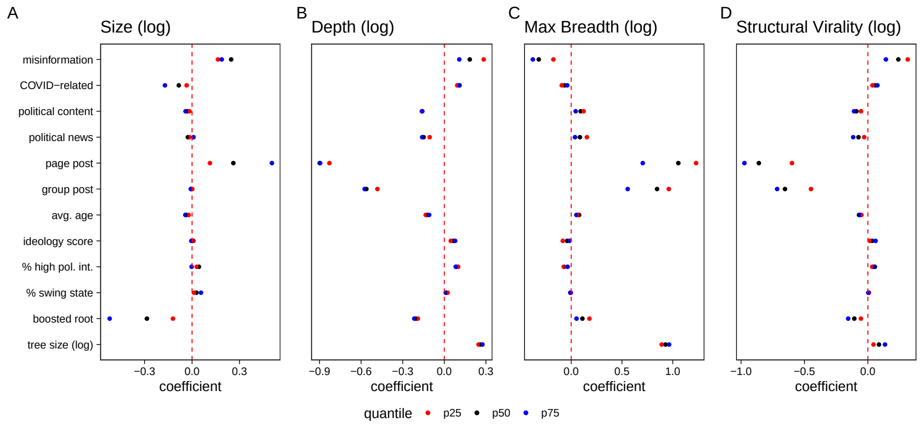

Figure S31: **Correlates of Diffusion Trees (Quantile Regression, Large Trees)** This figure is an extension to Figure 3 in the main paper. Quantile regression drops the assumption that variables operate the same at the lower and upper tails of the distribution as at the mean, thus helping identify potential differences across types of trees. The results of these regression models are consistent with the results reported in the main paper.

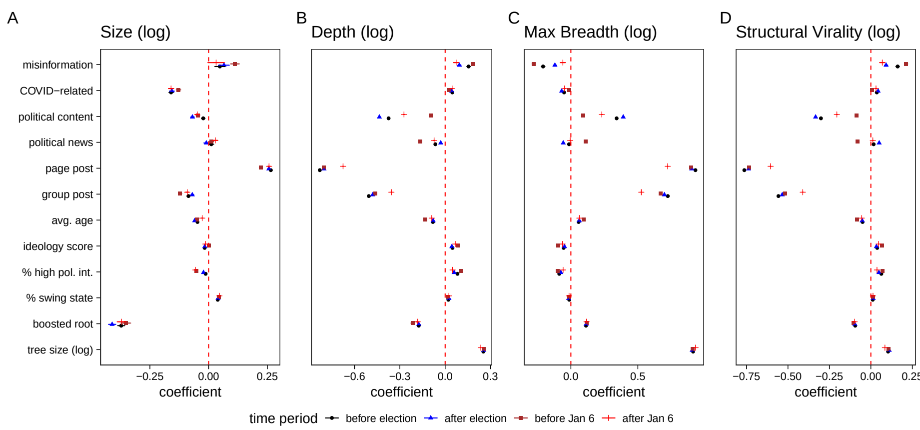

Figure S32: **Correlates of Diffusion Trees (Data Segmented by Time, Large Trees)** This figure is an extension to Figure 3 in the main paper and to S31. We split the data in four subsets: two weeks prior to November 3; two weeks post November 3; two weeks prior to January 6; and two weeks post January 4. The figure shows the coefficients for these different subsets.

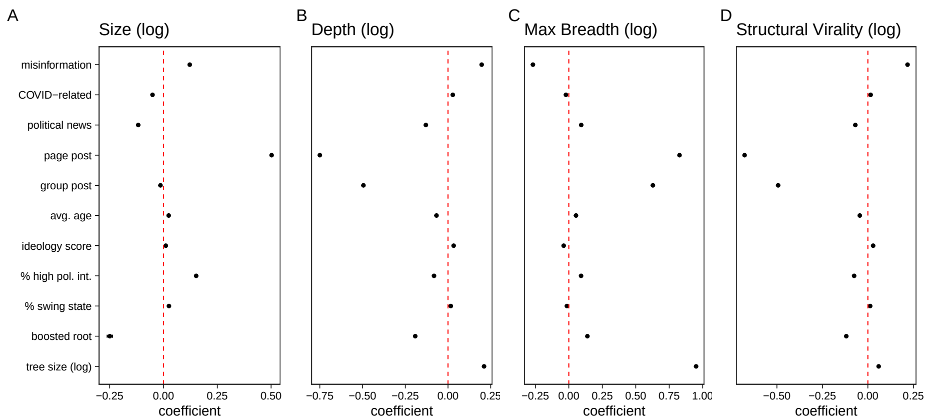

Figure S33: **Correlates of Diffusion Trees Classified as Political (Large Trees)** This figure is an extension to Figure 3 in the main paper: it subsets the data to just the trees that are classified as political. The results of these regression models are consistent with the results reported in the main paper.

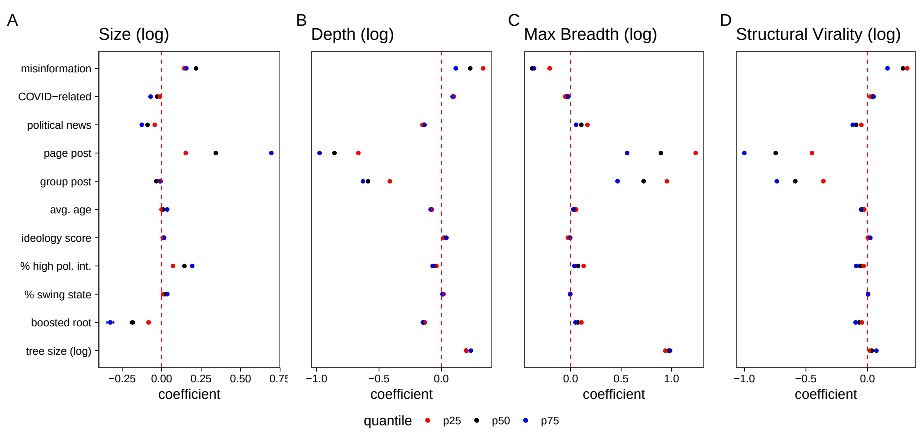

Figure S34: **Correlates of Diffusion Trees Classified as Political (Quantile Regression, Large Trees)** This figure is an extension to Figure 3 in the main paper and to Figure S33: it subsets the data to just the trees that are classified as political, and displays quantile regression estimates. Quantile regression drops the assumption that variables operate the same at the lower and upper tails of the distribution as at the mean, thus helping identify potential differences across types of trees. The results of these regression models are consistent with the results reported in the main paper.

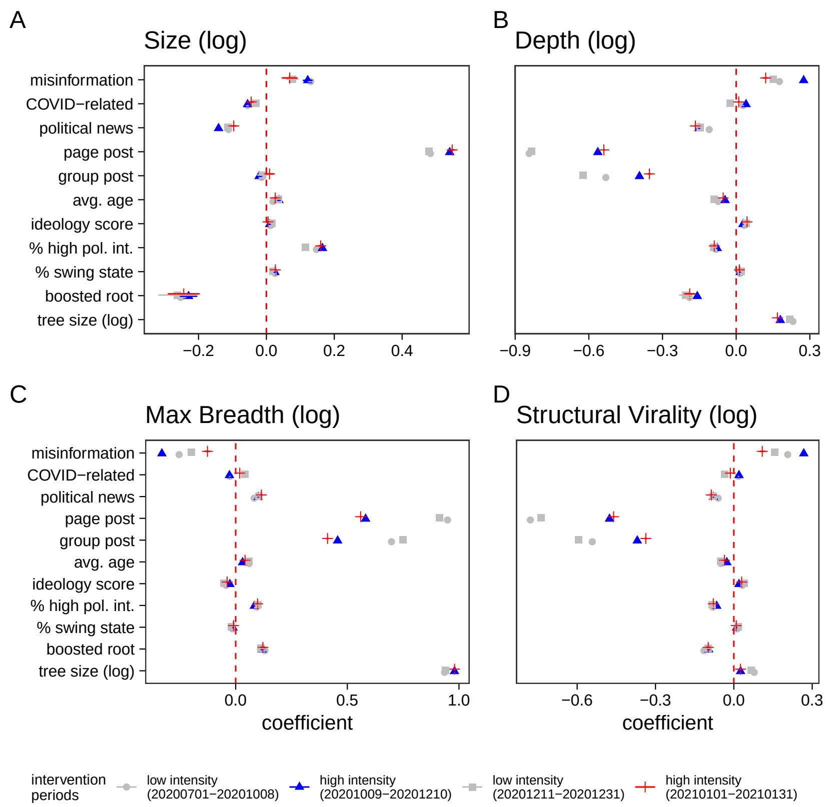

Figure S35: **Correlates of Diffusion Trees Classified as Political (Data Segmented by Intervention Period, Large Trees)** This figure is an extension to Figure 3 in the main paper. Here we split the data in the four intervention periods discussed in the main text but now the data only considers political trees.

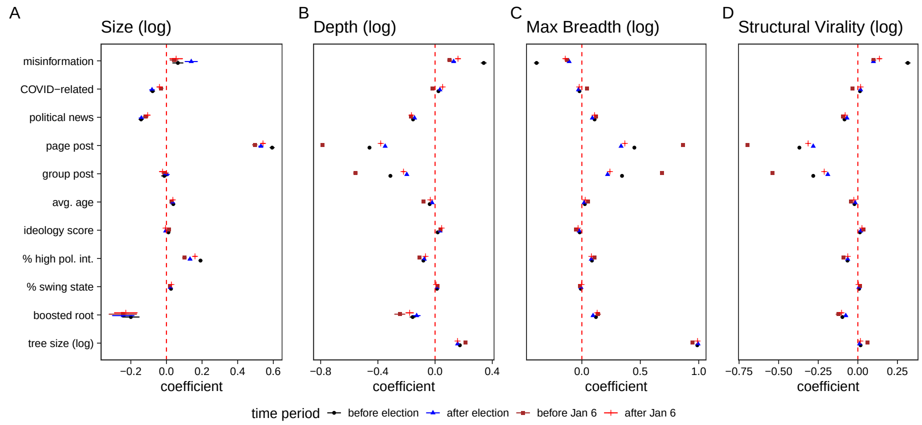

Figure S36: **Correlates of Diffusion Trees Classified as Political (Data Segmented by Time, Large Trees)** This figure is an extension to figures S33 and S34. The results are qualitatively similar.

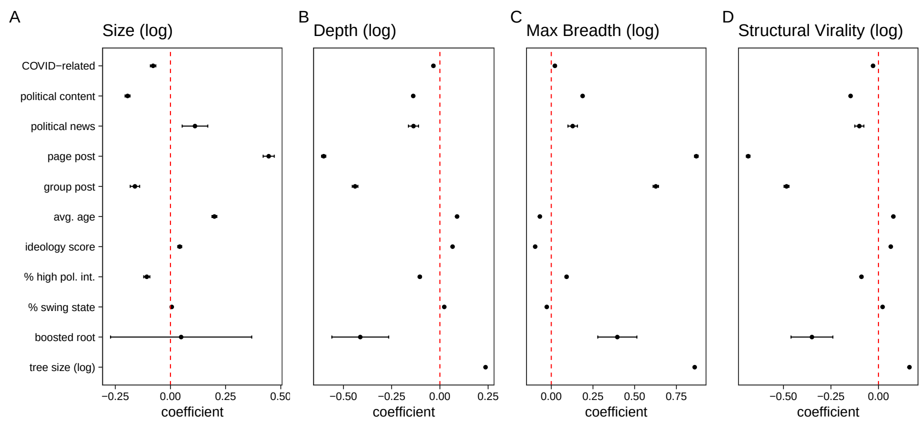

Figure S37: **Correlates of Diffusion Trees Classified as Misinformation (Large Trees)** This figure is an extension to Figure 3 in the main paper: it subsets the data to just the trees that are classified as misinformation. Other than when posted by Pages, misinformation trees grow larger when older users are involved, and when the content being diffused is classified as political news (a subset of political content). Political news diffuse mostly via broadcasting, hence the positive impact of this covariate on breadth (and negative impact on depth and structural virality).

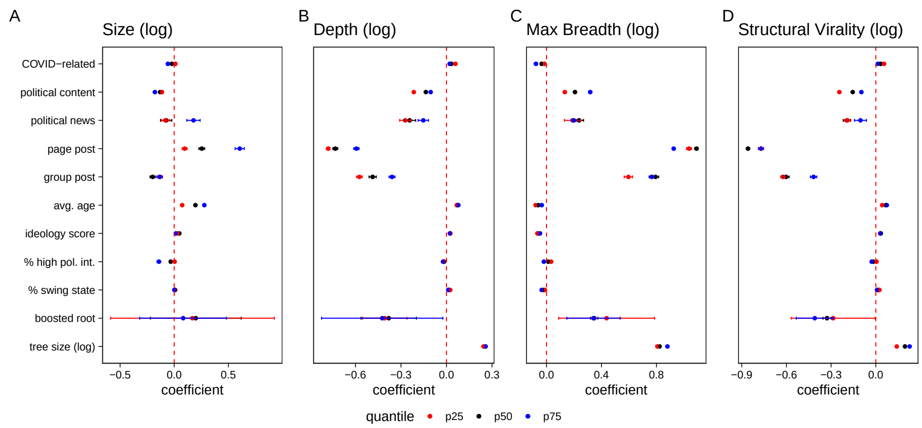

Figure S38: **Correlates of Diffusion Trees Classified as Misinformation (Quantile Regression, Large Trees)** This figure is an extension to Figure 3 in the main paper and to Figure S37: it subsets the data to just the trees that are classified as misinformation, and displays quantile regression estimates. Quantile regression drops the assumption that variables operate the same at the lower and upper tails of the distribution as at the mean, thus helping identify potential differences across types of trees. In this case, the positive effect of news content on tree size is driven mostly by the largest trees.

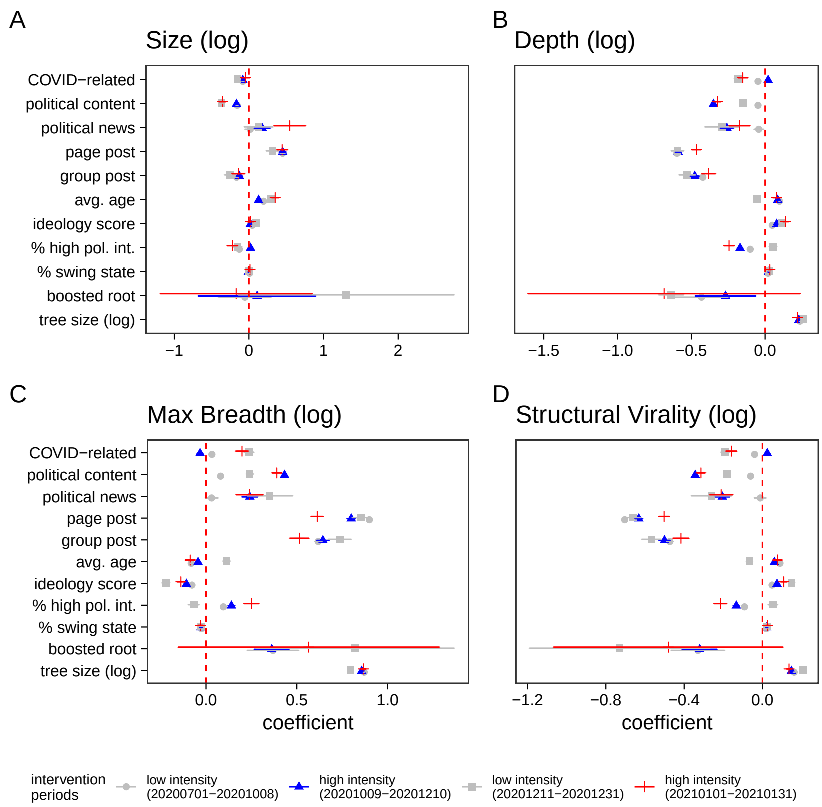

Figure S39: **Correlates of Diffusion Trees Classified as Misinformation (Data Segmented by Intervention Period, Large Trees)** This figure is an extension to Figure 3 in the main paper. Here we split the data in the four intervention periods discussed in the main text but now the data only considers misinformation trees.

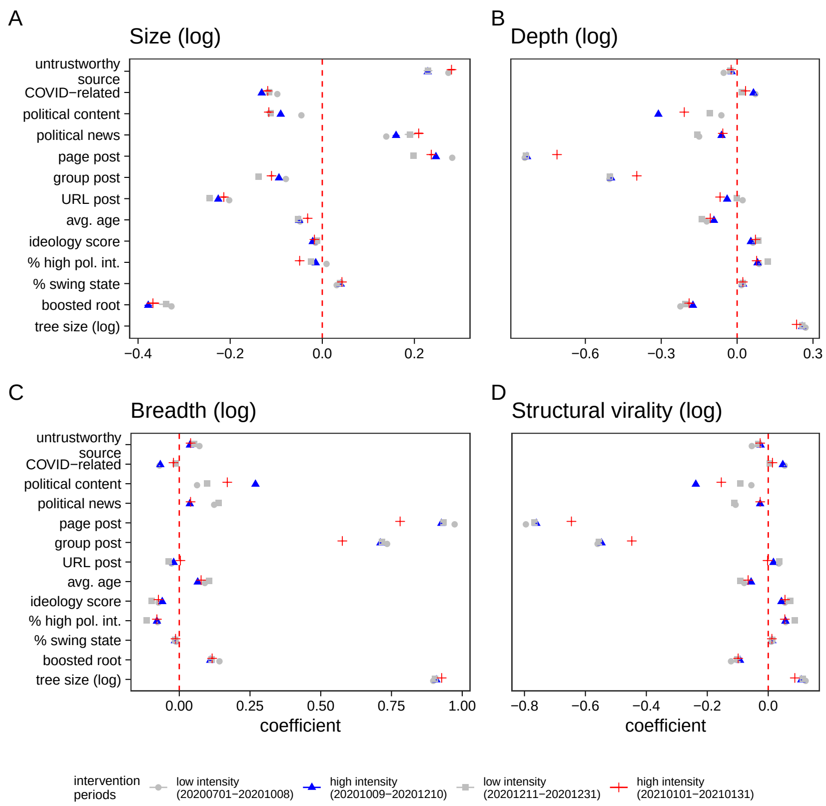

Figure S40: **Correlates of Diffusion Trees Classified as Untrustworthy (Large Trees)** This figure is an extension to Figure 3 in the main paper. Instead of sub-setting the the data to the trees classified as 'misinformation', here we look at the trees classified as 'untrustworthy'. Untrustworthy trees grow larger through broadcasting, consistent with the prevalence of Pages as root initiators in this category of trees, as discussed in sections S3.4 and S5.3.

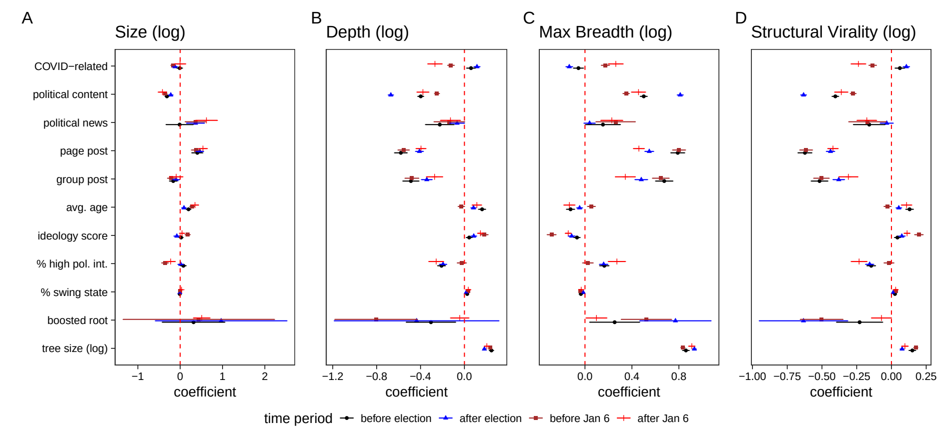

Figure S41: **Correlates of Diffusion Trees Classified as Misinformation (Data Segmented by Time, Large Trees)** This figure is an extension to figures S37 and S38.

Table S14: OLS regressions predicting tree-level metrics (all large trees)

Model:	Size (1)	Depth (2)	Breadth (3)	Virality (4)
Variables				
misinformation	0.1294*	0.1849*	-0.2613*	0.2122*
	(0.0025)	(0.0013)	(0.0015)	(0.0012)
COVID-related	-0.1250*	0.0781*	-0.0725*	0.0557*
	(0.0011)	(0.0005)	(0.0005)	(0.0004)
political content	-0.0149*	-0.1560*	0.1451*	-0.1288*
	(0.0010)	(0.0005)	(0.0005)	(0.0004)
political news	0.0076*	-0.1240*	0.0756*	-0.0563*
	(0.0013)	(0.0006)	(0.0005)	(0.0004)
page post	0.2737*	-0.8256*	0.9359*	-0.7664*
	(0.0007)	(0.0003)	(0.0003)	(0.0003)
group post	-0.0686*	-0.4857*	0.7056*	-0.5410*
	(0.0009)	(0.0005)	(0.0004)	(0.0004)
avg. age	-0.0461*	-0.1098*	0.0812*	-0.0709*
	(0.0005)	(0.0002)	(0.0002)	(0.0002)
ideology score	-0.0076*	0.0612*	-0.0682*	0.0504*
	(0.0003)	(0.0002)	(0.0001)	(0.0001)
% high pol. int.	-0.0089*	0.0832*	-0.0775*	0.0574*
	(0.0005)	(0.0003)	(0.0002)	(0.0002)
% swing state	0.0359*	0.0200*	-0.0130*	0.0114*
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
boosted root	-0.3520*	-0.2054*	0.1295*	-0.1119*
	(0.0020)	(0.0012)	(0.0007)	(0.0007)
tree size (log)		0.2612*	0.9062*	0.1126*
		(0.0001)	(0.0001)	(0.0001)
Fixed-effects				
Date	Yes	Yes	Yes	Yes
Fit statistics				
Observations	12,066,118	12,066,118	12,066,118	12,066,118
R^2	0.02480	0.48426	0.86139	0.51152
Within R^2	0.02448	0.47481	0.86073	0.50357

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: *: 0.01*

Table S15: OLS regressions predicting tree-level metrics (large political trees)

Model:	Size (1)	Depth (2)	Breadth (3)	Virality (4)
Variables				
misinformation	0.1220*	0.1970*	-0.2682*	0.2180*
	(0.0029)	(0.0017)	(0.0020)	(0.0016)
COVID-related	-0.0509*	0.0280*	-0.0217*	0.0152*
	(0.0013)	(0.0006)	(0.0006)	(0.0005)
political news	-0.1173*	-0.1296*	0.0927*	-0.0689*
	(0.0015)	(0.0006)	(0.0005)	(0.0004)
page post	0.5025*	-0.7499*	0.8270*	-0.6776*
	(0.0014)	(0.0008)	(0.0008)	(0.0006)
group post	-0.0139*	-0.4948*	0.6279*	-0.4932*
	(0.0020)	(0.0013)	(0.0012)	(0.0010)
avg. age	0.0242*	-0.0673*	0.0537*	-0.0445*
	(0.0010)	(0.0005)	(0.0004)	(0.0004)
ideology score	0.0107*	0.0331*	-0.0384*	0.0290*
	(0.0005)	(0.0002)	(0.0002)	(0.0002)
% high pol. int.	0.1519*	-0.0819*	0.0915*	-0.0756*
	(0.0011)	(0.0007)	(0.0006)	(0.0005)
% swing state	0.0254*	0.0163*	-0.0144*	0.0127*
	(0.0005)	(0.0003)	(0.0003)	(0.0002)
boosted root	-0.2493*	-0.1917*	0.1388*	-0.1187*
	(0.0063)	(0.0029)	(0.0015)	(0.0014)
tree size (log)		0.2110*	0.9507*	0.0593*
		(0.0003)	(0.0003)	(0.0002)
Fixed-effects				
Date	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,634,690	2,634,690	2,634,690	2,634,690
R^2	0.06780	0.52934	0.89178	0.57745
Within R^2	0.06619	0.46999	0.88920	0.53442

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: *: 0.01*

Table S16: OLS regressions predicting tree-level metrics (large misinformation trees)

Model:	Size (1)	Depth (2)	Breadth (3)	Virality (4)
Variables				
COVID-related	-0.0787*	-0.0336*	0.0215*	-0.0284*
	(0.0063)	(0.0034)	(0.0039)	(0.0031)
political content	-0.1944*	-0.1382*	0.1879*	-0.1465*
	(0.0057)	(0.0029)	(0.0034)	(0.0027)
political news	0.1111*	-0.1368*	0.1283*	-0.1008*
	(0.0295)	(0.0133)	(0.0145)	(0.0122)
page post	0.4453*	-0.6022*	0.8698*	-0.6852*
	(0.0129)	(0.0055)	(0.0057)	(0.0049)
group post	-0.1610*	-0.4395*	0.6268*	-0.4835*
	(0.0109)	(0.0074)	(0.0082)	(0.0066)
avg. age	0.1991*	0.0890*	-0.0684*	0.0784*
	(0.0056)	(0.0028)	(0.0031)	(0.0025)
ideology score	0.0420*	0.0657*	-0.0966*	0.0648*
	(0.0045)	(0.0021)	(0.0023)	(0.0019)
% high pol. int.	-0.1073*	-0.1039*	0.0919*	-0.0894*
	(0.0071)	(0.0036)	(0.0040)	(0.0032)
% swing state	0.0058	0.0227*	-0.0271*	0.0218*
	(0.0024)	(0.0014)	(0.0016)	(0.0012)
boosted root	0.0484	-0.4126*	0.3960*	-0.3499*
	(0.1632)	(0.0751)	(0.0598)	(0.0561)
tree size (log)		0.2367*	0.8600*	0.1630*
		(0.0016)	(0.0020)	(0.0018)
Fixed-effects				
Date	Yes	Yes	Yes	Yes
Fit statistics				
Observations	114,255	114,255	114,255	114,255
R^2	0.06616	0.41903	0.74845	0.41900
Within R^2	0.05284	0.34408	0.73752	0.34562

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: *: 0.01*

S5.15 Concentration of Shares

The platform data we analyze is aggregated at the tree level, so it is not possible to determine how many of the trees we analyze are generated by the same users, i.e., misinformation trees could repeatedly arise from the same small group of individuals, thus limiting their reach to a narrow segment of the population. To get a sense of what fraction of the total user base is responsible for the diffusion dynamics we analyze, we used the panel data described in section S2.2 to calculate the cumulative distribution function (CDF) for original posts, re-shares, and exposures. We show these CDFs in Figure S42. In order to address some of the limitations of our participant sample, which was not representative of the U.S. Facebook adult population, we also report a weighted version of the CDF in figure S43. As the figures show, a small minority of users create most of the posts and re-shares, especially for content classified as misinformation.

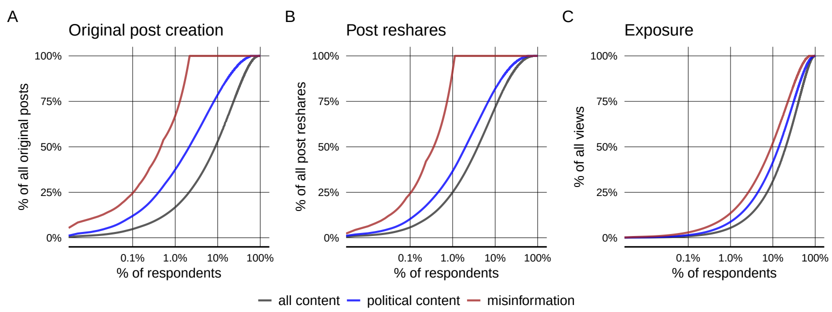

Figure S42: **Concentration in Posting, Re-sharing, and Exposures (Unweighted Panel Data)** A small minority of users create most of the posts and re-shares, especially for content classified as misinformation.

S5.16 Reach over Time (Large Trees)

In Figures S44, S45, S46 and S47 and table S17 we show additional findings that contextualize the results reported in Figure 4 of the main text. Pages attain substantially higher reach (Figures S44 and S45 and table S17). Misinformation trees accumulate fewer views than political trees and all trees during most of the observation window (Figure S46). While trees initiated by Page posts receive more views, on average, than Group posts and user posts, in the aggregate, misinformation cascades triggered by user posts are viewed the most on any given day (see Figure S47). As we show in Figures S42 and S43, we estimate the fraction of users posting and re-sharing misinformation to be as small as $\sim 1\%$ of all US-based users.

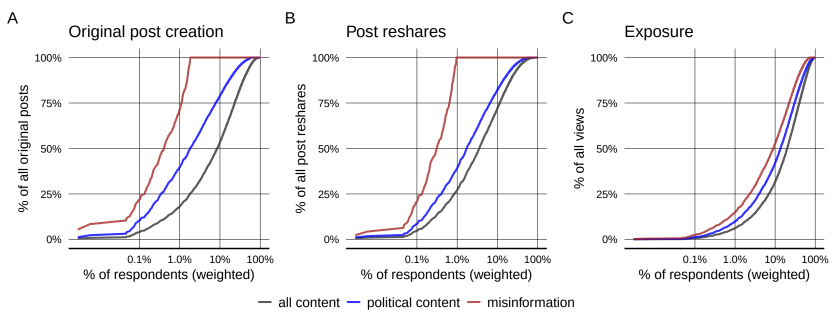

Figure S43: **Concentration in Posting, Re-sharing, and Exposures (Weighted Panel Data)** Adding weights to the panel data (to make it representative of the US population, see S2.2) does not change the patterns identified in S42.

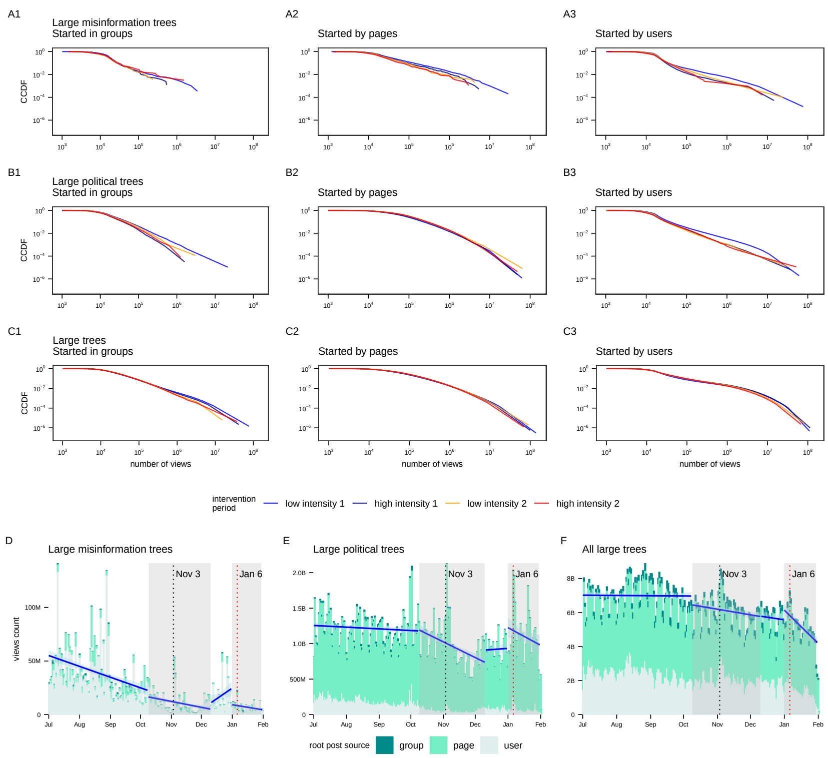

Figure S44: **Distribution of Views for Content Posted across Intervention Periods** This figure is an extension of Figure 4 in the main text. (A1-C3) Log-log version of the view distributions across intervention periods. (D-F) Changes in the reach of large trees, where blue lines are linear trends, fitted separately for each of the four intervention periods.

Content	Subset	p5	p50	p95	Avg	D	p
			Full dataset, Groups				
Misinformation	low intensity 1	4229.6	13848	43694.80	25692.9		
	high intensity 1	3279.3	12163	40740.70	17773.6	0.10	<1e-6
	low intensity 2	3243.85	10920	41912.00	16342.6	0.13	<1e-6
	high intensity 2	3117.7	12896	46119.70	22987.9	0.11	<1e-6
Political	low intensity 1	4441.8	14521	86125.60	29377.3		
	high intensity 1	3638	12329	60619.75	19980	0.09	<1e-6
	low intensity 2	4070.75	13387.5	78010.50	24459.6	0.05	6.7e-16
	high intensity 2	3781.5	12412	66042.50	21450.2	0.09	<1e-6
All trees	low intensity 1	5186	17042	125796.40	46584		
	high intensity 1	5122	17824	132908.65	45395.4	0.03	<1e-6
	low intensity 2	5503	18348	131528.20	42855.8	0.04	<1e-6
	high intensity 2	5329	17620	125871.75	40654.6	0.02	<1e-6
			Full dataset, Pages				
Content	Subset	p5	p50	p95	Avg	D	p
Misinformation	low intensity 1	5584	19943	311103.90	115065.2		
	high intensity 1	4866	18184	194699.00	65600	0.06	4.7e-05
	low intensity 2	4870.2	17469	143919.20	54913.4	0.09	<1e-6
	high intensity 2	4227.9	15453.5	111828.05	43487.4	0.15	1e-15
Political	low intensity 1	7101	41139	400810.80	117229.1		
	high intensity 1	6953	45708	453646.00	127983.2	0.04	<1e-6
	low intensity 2	7675	52782.5	479665.85	141855.8	0.08	<1e-6
	high intensity 2	7368	53758	494840.95	141227.1	0.08	<1e-6
All trees	low intensity 1	6912	36246	432106.80	126034.7		
	high intensity 1	6846	39032	479461.20	133003.8	0.03	<1e-6
	low intensity 2	7090.25	41580	477201.75	135866.7	0.04	<1e-6
	high intensity 2	7231	42907	473496.00	131549.8	0.05	<1e-6
			Full dataset, Users				
Content	Subset	p5	p50	p95	Avg	D	p
Misinformation	low intensity 1	6650.7	17917	63139.20	49059.5		
	high intensity 1	5274.45	16643	41677.80	28421.3	0.11	<1e-6
	low intensity 2	5930.5	17193	56084.00	34002.4	0.05	<1e-6
	high intensity 2	5024.9	15548	51550.10	25992.9	0.17	<1e-6
Political	low intensity 1	5535	15941	66242.80	36251.8		
	high intensity 1	4342	13173	50757.10	23485.5	0.12	<1e-6
	low intensity 2	4880	14543	46995.00	22790.9	0.09	<1e-6
	high intensity 2	4347	12062	53881.00	23099.9	0.16	<1e-6
All trees	low intensity 1	6009	17464	175530.80	111372.7		
	high intensity 1	5417	17380	254493.90	126594	0.04	<1e-6
	low intensity 2	5617	17216	246870.65	110917.5	0.03	<1e-6
	high intensity 2	5155	16687	230998.70	96920.2	0.05	<1e-6

Table S17: **Views for all large trees initiated by Groups, Pages, and Users across intervention periods.** Summary statistics (5th, 50th, and 95th percentiles and average) for views distributions, and D statistic and p-value for KS tests comparing the distributions to the baseline of the first period of low intensity interventions.

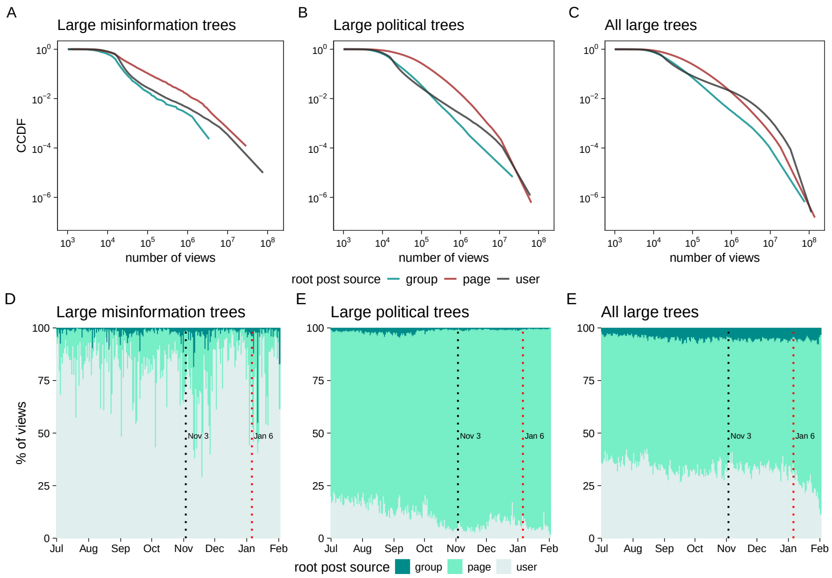

Figure S45: **Distribution of Views for Content Posted by Users, Pages, Groups** These CCDFs suggest that, even if user posts represent a majority of all large cascades, overall they do not accumulate as many views as the posts published by Pages, especially for misinformation and political content: in these groups (A-B) there is a clear over-representation of Page posts. Posts published in Groups are viewed, on average, by significantly fewer people than posts from users and Pages.

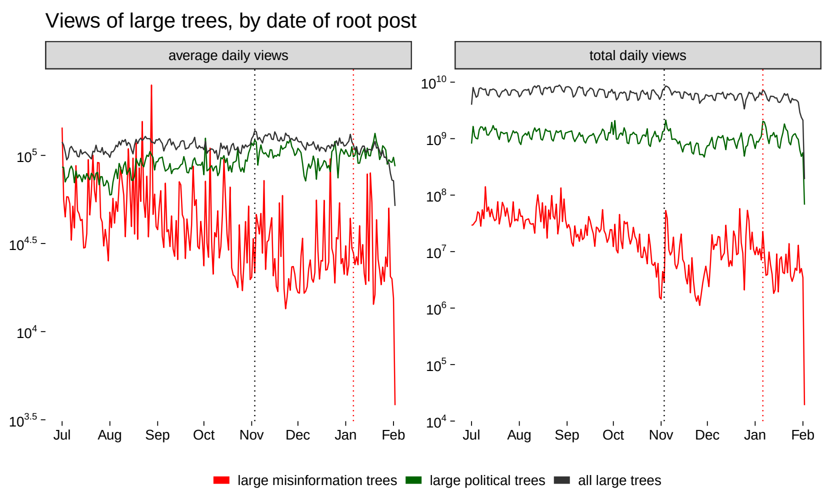

Figure S46: **Average and Total Daily Views for Large Trees** This figure is an extension to Figure 4 in the main text, casting light on temporal changes in the reach of large trees. Overall, misinformation trees receive fewer views than political trees or all trees.

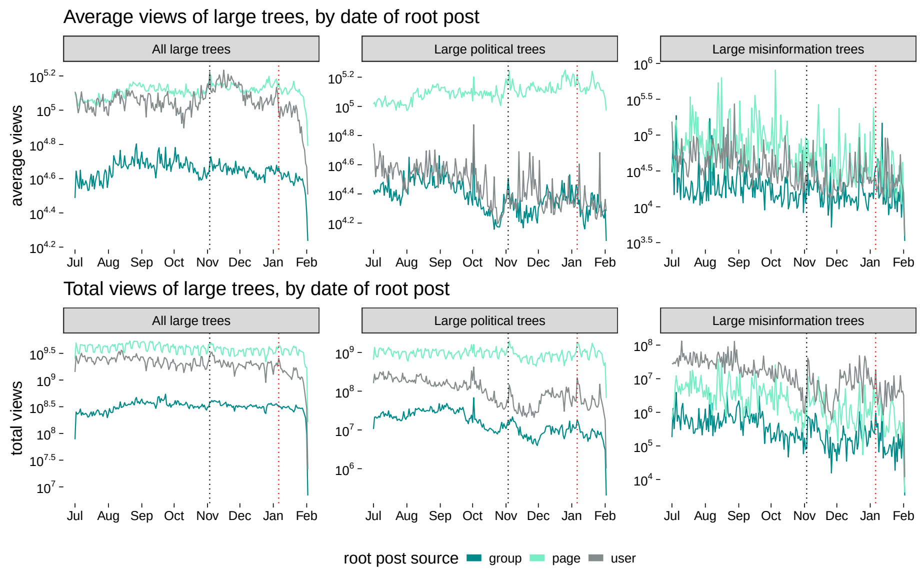

Figure S47: **Average and Total Daily Views** This figure is an extension to Figure 4 in the main text and Figure S46, adding information about the root post source for all large trees, political trees, and misinformation. Misinformation cascades triggered by user posts are, on any given day, viewed the most.

S6 Supplementary Results: Small Trees

The diffusion trees we analyze in the main text and in S5 (i.e., those with 100+ re-shares) represent 1.2% of all trees with at least 1 re-share. In this section, we focus on the small trees (i.e., the 98.8% trees with less than 100 re-shares and accumulating 45.4% of all views).

S6.1 Structural Differences (Small Trees)

In Figure S48 we summarize the structural properties of diffusion trees below the 100 re-shares threshold to explore the potential for bias resulting from the selection of this cutting point. The figure displays the average tree depth, breadth, and virality for trees of all sizes for different content categories (misinformation, political, all) and for different root post source (users, Pages, Groups). The vertical, dashed line identifies the threshold used to identify large trees (note that the x-axis is truncated at 1000 re-shares to improve the visualization of data points below the threshold). The plots show that our main conclusions (i.e., misinformation trees are deeper but less broad; also trees initiated by users) hold below the threshold, i.e., they are also true for the category of trees labeled as "small". For the much smaller trees with $k < 10$ re-shares the differences between the categories become smaller.

S6.2 Reach Distribution (Small Trees)

In Figure S49 we show the average reach distribution for the set of small trees, grouped by type of initiating account (Groups, Pages, users) and by type of content (misinformation, political). The graphs show that (a) Pages and to a lesser extent Groups accumulate, on average, higher reach with smaller trees; (b) misinformation trees garner, on average, less views; and (c) whether misinformation posts are published by Groups, Pages or users does not really matter for this subset of small trees, in terms of increasing reach; the smallest trees ($k < 10$ re-shares) accumulate more average views when published by Groups.

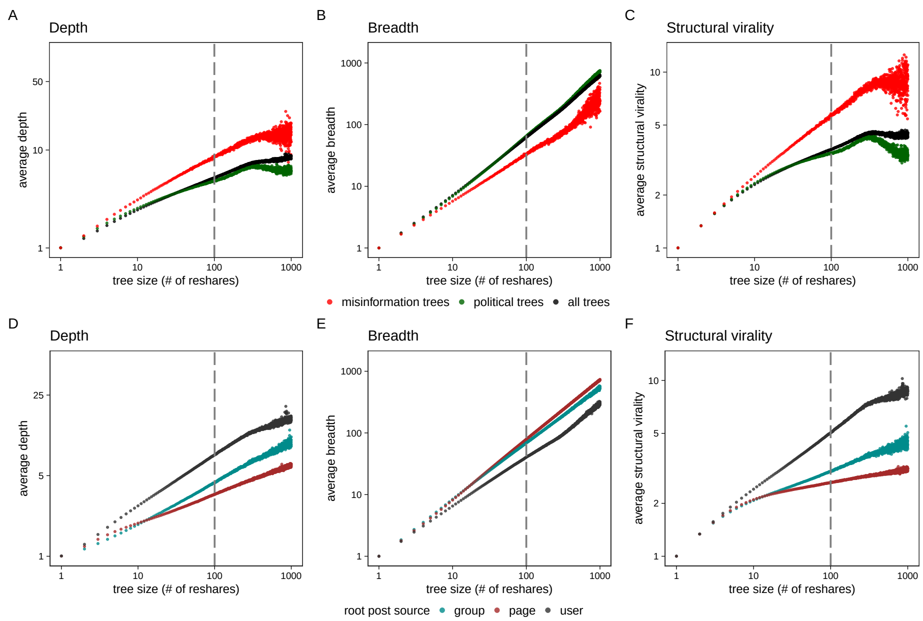

Figure S48: **Structural Properties of Trees Below and Above the $k = 100$ Size Threshold.** This figure displays the average tree depth, breadth, and structural virality for trees below the $k = 100$ re-share threshold and right above it (up to $k = 1000$ re-shares). The conclusions we report in the main text of the paper for large trees also hold for small trees.

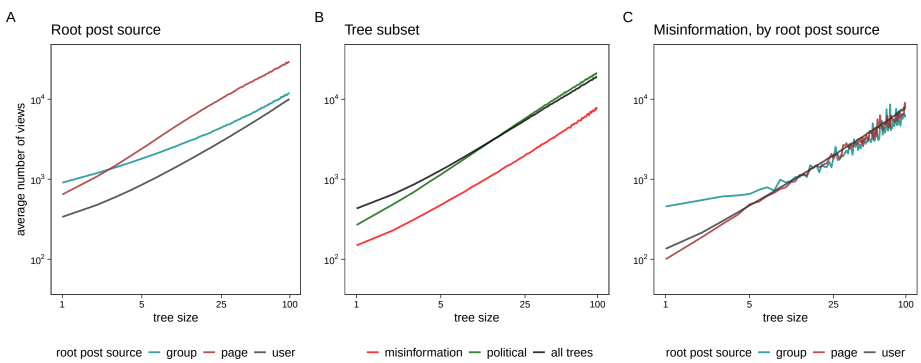

Figure S49: **Reach Distribution of Content in Small Trees** This figure is an extension to Figure 4 in the main text, casting light on the average reach distribution for trees with less than 100 re-shares (see section S3.1 for the rationale of this threshold). Overall, Pages accumulate more views than Groups and users with diffusion trees of the same size, but not when the content diffused is labelled as misinformation which, on average, reaches less people regardless of the original poster (with the exception of the smallest trees).

S7 Pre-Analysis Plan

The PAP included in the next section was registered with the OSF in 2020, before the start of data collection (see section S1.2 for more details).

Diffusion and Virality of Political Content

0. Control Rights and Auditing

The lead authors of this study are Sandra González-Bailón (Penn) and David Lazer (Northeastern). They retain final discretion over what is reported in the paper and how it is reported. In the event of disagreement within the research team about the appropriate presentation or interpretation of the results, the final decision will rest with them.

Facebook researchers Pablo Barbera and Winter Mason are Facebook co-authors of the study. Talia Stroud and Joshua A. Tucker are lead researchers of the FB 2020 Election Research Study, of which this PAP is a part, and will serve as co-last authors on the paper.

1. Introduction

1.1. Summary

The main goal of this paper is to determine how Facebook facilitated the spread of political content in the months leading up to the 2020 Election. This research will quantify diffusion dynamics, differentiate broadcasting from viral spreading, identify the type of content that triggered broader diffusion, and measure how different platform affordances (pages/groups/re-shares) shape those diffusion dynamics. This research will complement the reshare holdout experiment by offering behavioral indicators of content diffusion on Facebook and identifying the mechanisms behind those diffusion dynamics, which result in exposure to political information among users who might not have been exposed to that content otherwise.

1.2. Motivation

Social media have made reaching wider audiences much easier: their mechanisms for sharing content often results in chain reactions that spread beyond the circles for which the content was originally intended. These chain reactions can arise organically or they can be engineered through seeding strategies (e.g., sponsored posts). This project will quantify whether political and non-political content spread in the same way among Facebook users (US adults); determine if certain types of content (e.g., unreliable news) are more likely to spread among certain populations (e.g. older users); and assess if diffusion is more likely to unfold among ideologically similar users. Ultimately, this research has two goals:

- Determine if certain populations are more vulnerable to misinformation and echo-chamber effects; and
- Determine if there is an asymmetry in the information diffusion patterns triggered by right-leaning and left-leaning actors.

1.3. Prior Work

Research on information diffusion is extensive, but the number of studies retracing (and characterizing) the chain reactions underlying diffusion events are much smaller. One early example analyzed the propagation of massively circulated Internet chain letters [1]. This study employed a data-processing method to transform mailing-list archives into diffusion trees where nodes were signatories of online petitions and connections mapped public re-posting of the petition. For a petition that accumulated 20,000 distinct signatories, for instance, the analyses show that the diffusion did not fan out, reaching most people through a few steps, but instead progressed through long and narrow paths. This suggests that successful diffusion events become successful through the impact that many people have on a small number of others (at least in the context of political petitions). A later study scaled the analyses up to consider diffusion on Twitter [2]. This study analyzed 622 million unique pieces of content (1.2 billion “adoptions”, given that the same content was posted by multiple users independently). One of the main takeaways of this study is that large diffusion events are the exception: only 1 out of 4,000 cascades involved at least 100 nodes. This reduced the number of diffusion events to ~ 220,000 (out of the 1.2 billion). This study also introduced the measure of “structural virality”, which helps quantify the heterogeneity of information cascades. This measure was adopted in a more recent study that determined that false information spreads more (farther, faster, and more broadly) than true information on Twitter [3]. The study analyzed ~ 126,000 cascades with verifiable content (e.g., links to the websites of fact checking organizations, ~24,000 cascades were true, 83,000 were false). Past work has also found that the propagation of false content on Facebook takes place through cascades that run deeper in the social network than reshare cascades on other topics [4]. Here we will build on this prior work and characterize cascades as diffusion trees with quantifiable measures of virality. The goal is to determine pathways of diffusion for different types of content and different types of users.

2. Research Design

2.1. Study Overview

This project aims to (a) identify the types of content that are more likely to reach virality; (b) determine whether some populations are more susceptible to the spread of unreliable information; (c) assess if the diffusion of political content follows asymmetric patterns, with significant differences between the right and the left; and (d) determine how platform affordances (pages/groups/re-shares) contribute to the spread of content. The analyses will focus on diffusion dynamics and, in particular, differences in the diffusion trees that arise as content percolates through Facebook. The basic unit of analysis will be diffusion trees. A diffusion tree is a branching structure where nodes are “actors” posting/sharing content, and branches encode the sharing activity of that content. Diffusion trees map information cascades (‘trees’, ‘cascades’, and ‘diffusion events’ are used interchangeably in this paper). The analysis will compare the structural properties of those trees for (1) different types of content and (2) different types of users. Every tree corresponds to an independent introduction of a piece of content. This means that there will be as many diffusion events as independently introduced posts: the same piece of content (e.g., URL) can be posted by multiple users but they will count as independent cascades if they are not part of the same diffusion event.

2.2. Research Questions

- What type of content is more likely to trigger viral diffusion?
    - Does sponsored content spread more/less than organic content?
    - Does political content spread more/less than non-political content?
    - Does misinformation about covid spread more/less than reliable information?
    - Does political misinformation spread more/less than reliable political information?
    - Do URLs to reliable news sources spread more/less than URLs to non-reliable sources?
    - Do political news spread more/less than non-political news?

- Does content diffuse differentially across demographic groups?
    - What type of content spreads across ideological divides?
    - What type of content spreads among users with high political interest?
    - What type of content spreads among users with low political interest?
    - Do older users contribute more to the diffusion of misinformation?
    - Is there an asymmetry in information diffusion dynamics between left-leaning and right-leaning users?
    - Are the differences in the type of content that spreads in swing states versus other states?

2.3. Data

The key unit of analysis in this paper will be "shares", the basic building block of diffusion trees. We will analyze data aggregated in two Data Frames (full definitions of variables below).

Descriptive statistics drawn from Data Frame 1 will allow us to identify the threshold to select the diffusion trees in Data Frame 2 (e.g., content reaching a minimum number of views). Data will be collected from September 1 to February 1, with preliminary analyses conducted half-way through the observation window to make sure data processing produces the right measurements. These Data Frames summarize the data on which this study will be based; they are not output tables but a description of the variables that we will use to produce the results outlined in sections 3 and 4.

- Data Frame 1: content diffused.
    - Variables: <original_content_id>;<parent_content_id>; <content_id>; <content_audience_ideology>; <content_type>; <content_type_civic>; <content_type_civic_misinformation>; <content_type_covid>; <content_type_covid_misinformation>; <content_type_origin>; <first_time_posted>; <first_time_shared>; <last_time_posted>; <last_time_shared>; <shares>; <views>.

- Data Frame 2: diffusion trees (subset of Data Frame 1, depending on threshold of minimum views).
    - Variables: <original_content_id>; <tree_size>; <tree_depth>; <tree_breadth>; <structural _virality>; <time_to_depth>; <time_to_size>; <%_Conservative>; <%_Liberal>; <ideological_score>; <age_avg>; <age_sd>; <%_users_high_political_interest>; <%_pages>; <%_sponsored_root>; <%_battleground_state>.

Definitions:

- <%_users_high_political_interest>: fraction of users in the tree that are classified as having high political interest based on their on-platform civic engagement.
- <%_Liberal>: fraction of users/pages in the tree that are classified as liberal (for pages, we use audience ideology).
- <%_Conservative>: fraction of users/pages in the tree that are classified as (for pages, we use audience ideology) conservative.
- <%_pages>: fraction of diffusion event in trees that are pages.
- <%_sponsored_root>: fraction of views in root node of the tree that take place via a sponsored vs organic content view.
- <age_avg>: average age of users in the tree.
- <age_sd>: standard deviation of users age in the tree.
- <%_battleground_state>: fraction of users in the tree that are located in a swing state[1].
- <content_id>: unique identifier for every piece of content posted independently (the id of the same content will be different if different users post it separately). This is the key connecting the two Data Frames (i.e., it will allow us to access information on content type when analyzing trees).
- <content_audience_ideology>: average predicted ideology of users who follow a Page (for Page posts) or are members of a group (for Group posts). Missing for user posts.
- <content_type>: (for URLs only) whether URL is from a domain identified as civic news by Facebook's internal classifier.
- <content_type_civic>: sub-category as defined by Facebook's topic classification; e.g. politics, health and medical, science and tech, business... These sub-topics overlap. For tree-level predictions, we will classify trees as civic if 50% or more nodes in the tree are classified as civic.
- <content_type_civic_misinformation>: binary, content rated "false" by the Third Party Fact-Checking Program (3PFC).
- <content_type_covid>: binary.

[1] Following the two most recent Electoral College Ratings by the Cook Political Report prior to August, we defined as battleground states those whose complete electoral geography was considered in the "Toss Up", "Lean Democrat", or "Lean Republican" in at least one of the reports. The states that met this criterion are the following: (a) "Toss up" states: Arizona, Georgia, Maine, North Carolina (9.06% of the total US population); (b) "Lean Democrat" or "Lean Republican" states: Florida, Michigan, Minnesota, New Hampshire, Pennsylvania, Wisconsin, Iowa, Ohio, Texas (26.32% of the US total).

- <content_type_covid_misinformation>: binary, content rated “false” by the Third Party Fact-Checking Program (3PFC).
- <content_type_origin>: whether the post was published by user, group, or page.
- <first_time_posted>: time stamp of the first time content is posted.
- <first_time_shared>: time stamp of the first time content is shared.
- <ideological_score>: i_s = (C – L) / (C + L), it is 0 if there is an equal number of liberal and conservatives in the tree.
- <last_time_posted>: time stamp of the last time content is posted.
- <last_time_shared>: time stamp of the last time content is shared.
- <original_content_id>: for reshared content, content_id for the first post that initiated a diffusion tree.
- <parent_content_id>: ID necessary to reconstruct intermediate steps in resharing cascades (e.g., reshares more than one hop away from the source; in direct reshares of the original content this is equal to <original_content_id>.
- <shares>: cumulative number of shares.
- <structural _virality>: average distance *d* between all pairs of nodes in a diffusion tree *T*; for *n* > 1 nodes, $v(T) = \frac{1}{n(n-1)}\sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$, where d_{ij} stands for the length of the shortest path between nodes *i* and *j*.
- <time_to_depth>: time it takes to reach each depth (in minutes).
- <time_to_size>: time it takes to reach a certain size (in minutes).
- <tree_breadth>: maximum breadth (maximum number of nodes over all depths).
- <tree_depth>: number of edges from a node in the tree and the root node. Tree depth is the maximum depth of the nodes in the cascade.
- <tree_size>: number of unique users in the cascade.
- <views>: number of impressions.

2.4. Classifiers

Some of the variables included in Data Frames 1-2 require applying classifiers to the raw platform data. What follows is a brief description of how these classifiers operate (a full description of these classifiers can be found in the global document “List of Classifiers”):

- <u>Content type</u>: (for URLs only) whether URL is from a domain identified as civic news by Facebook’s internal classifier. Civic content is defined as content that is relating to politics or social issues. News content is defined as content that reports on current events and follows journalistic standards.

- <u>Covid-content</u>: whether URL is predicted to be related to the Coronavirus Disease 2019 (content that explicitly mentions COVID-19, health concepts related to the virus, policy or public health guidance or the economic impact of the virus, as well as people’s reactions to the situation), using Facebook’s internal classifier.

- <u>High political interest</u>: users in the top 20% of exposure to/engagement with civic content on Facebook. We will use two operationalizations for robustness test purposes: (1) a measure based on exposure to civic content on the platform; and (2) a measure based on engagement with civic content (comments, likes, shares).

Definition (2) will give us a more stringent measure of 'political interest'. We will also consider a measure that increases this threshold to the top 10%.

- Untrustworthy content: we will use two operationalizations for robustness checks. (1) URLs from domains with 2+ lifetime strikes for misinformation. A strike is defined as a post rated "False" by a third-party fact-checker or a post that matches another post rated "False" by a third-party fact-checker; and (2) content classified as "unreliable" if it is published by a source included in publicly available databases of unreliable sources (e.g., mediabiasfactcheck.com, factcheck.org). We will also analyze unlabeled content (i.e., missing data) and compare it to the reliable/unreliable categories.

- Misinformation: URLs rated "False" by a third-party fact-checker in the list that Facebook uses to make enforcement decisions on misinformation.

- User ideology: the labels "liberal" and "conservative" are assigned to users based on Facebook's internal ideology classifier. For additional details, see the gobal "List of Classifiers" document.

3. Descriptive Statistics

Figure shells with descriptive statistics for the content variables in Data Frame 1. In this section, we will also discuss missing data (i.e., unlabeled content).

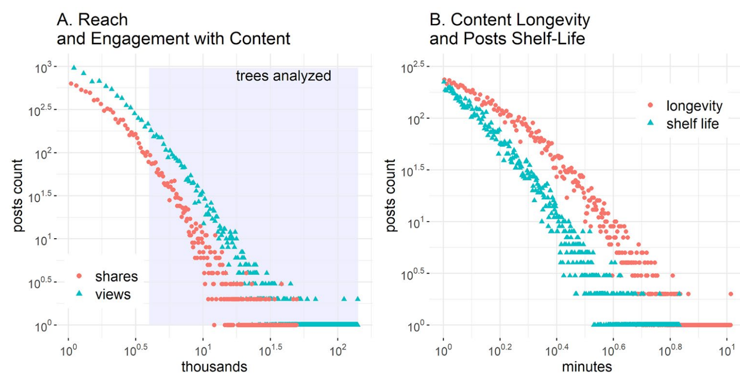

Figure 1: Descriptive Distributions for Reach and Engagement. Random synthetic data. Panel A will plot the distribution of posts in terms of total number of views and total number of shares. Diffusion trees will be built with a subset of these posts (those with a minimum number of views, shaded area). Panel B will plot the distribution of posts in terms of longevity (last_time_posted - first_time_posted) and 'shelf-life' (last_time_shared - first_time_shared).

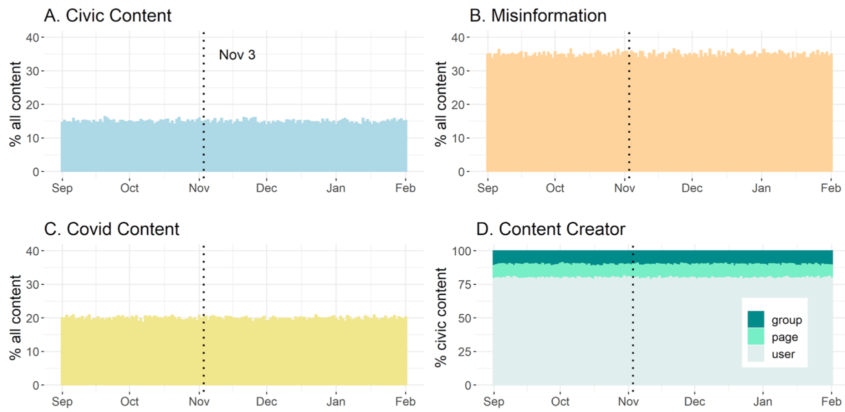

Figure 2: Content Volume over Time. Random synthetic data. This figure will plot the count of content classified as "civic" (A), "misinformation" (B), and "Covid-related" (C). Panel D plots the share of civic content by creator: individual user, group or page. In this random example, there are no temporal fluctuations -- but we expect the relative volume share to change over time.

4. Empirical Strategy

The analyses will focus on tests around four aspects of the data:

4.1. Diffusion Reach

- T1: test if content reach (number of users exposed who share the content) results from broadcasting (panel A) or from viral diffusion (panel B). Plot cumulative distributions of tree statistics in Data Frame 2 to determine which model (broadcasting vs virality) is more prevalent. We expect most cascades to fall between scenarios illustrated in panels A and B.

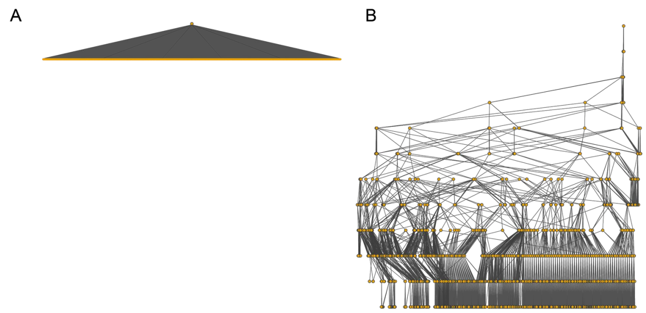

Figure 3: Schematic Representation of Broadcasting and Viral Trees. Synthetic data. Two diffusion events may involve the same number of unique users sharing content but differ greatly in how the content percolates through a network. In broadcasting events (panel A) all shares happen one step removed from the origin. In viral events (panel B) diffusion unfolds through nested layers of sharing activity. The measures of depth, breadth and structural virality in Data Frame 2 will allow us to quantify these different diffusion structures.

4.2. Misinformation

- T2: test if content labelled as 'misinformation' spreads differently compared to reliable content (for all content and within the categories in the <content_type_civic> variable). Use Kolmogorov-Smirnov tests to compare cumulative distributions of tree statistics in Data Frame 2 for these two types of content. Null hypothesis: there are no differences in spreading patterns of reliable and unreliable content.

- T3: same as T2 but for covid content only (for all content and within the categories in the <content_type_civic> variable).

4.3. Content Origin

- T4: test if content posted by users, groups, or pages spread differently (for all content and within the categories in the <content_type_civic> variable). Use Kolmogorov-Smirnov tests to compare cumulative distributions of tree statistics in Data Frame 2 for these three types of content. Null hypothesis: there are no differences in spreading patterns regardless of identity of root account.

- T5: test if content classified as having a majority of 'sponsored' views spreads differently from organic content (for all content and within the categories in the <content_type_civic> variable). Use Kolmogorov-Smirnov tests to compare cumulative distributions of tree statistics in Data Frame 2 for these two types of content. Null hypothesis: there are no differences in spreading patterns of sponsored and organic content.

4.4. Demographics

- T6: test if the diffusion of misinformation (for all content and within the categories in the <content_type_civic> variable) is more prevalent among older users. Compare average age distribution for 'misinformation' trees with average age distribution for 'reliable content' trees. Null hypothesis: on average, users participating in the diffusion of misinformation are older. Use non-parametric test with bootstrapped CI.

- T7: same as T6 but for covid content only (for all content and within the categories in the <content_type_civic> variable).

- T8: test if the diffusion of misinformation (for all content and within the categories in the <content_type_civic> variable) is more prevalent among conservative users. Compare the share of conservative/liberal users participating in 'misinformation' trees with the share of conservative/liberal users participating in 'reliable content' trees. For every tree we have an ideological_score = (C – L) / (C + L)). Null hypothesis: on average, the diffusion of misinformation involves the same share of conservative and liberal users (i.e., ideological_score = 0). Use non-parametric test with bootstrapped CI.

- T9: same as T8 but for covid content only (for all content and within the categories in the <content_type_civic> variable).

- T10: test if diffusion patterns change across liberal/conservative users. Separate trees into two categories using the ideological composition of users: 'liberal user majority', 'conservative user majority'. Use Kolmogorov-Smirnov tests to compare cumulative distributions of tree statistics in Data Frame 2 for these two groups (for all content and within the categories in the <content_type_civic> variable). Null hypothesis: there is no difference in information diffusion dynamics for liberals and conservatives (i.e., there is no evidence of information asymmetry).

- T11: test if the diffusion of misinformation is less prevalent among users with high political interest. Compare the fraction of users with high political interest in 'misinformation' trees with the fraction of high political interest users in 'reliable content' trees (for all content and within the categories in the <content_type_civic> variable) . Null hypothesis: on average, the diffusion of misinformation involves a lower percentage of users with high political interest. Use non-parametric test with bootstrapped CI.

- T12: same as T11 but for covid content only (for all content and within the categories in the <content_type_civic> variable) .

5. Empirical Results

Figure shells with main results from tests T1-T12:

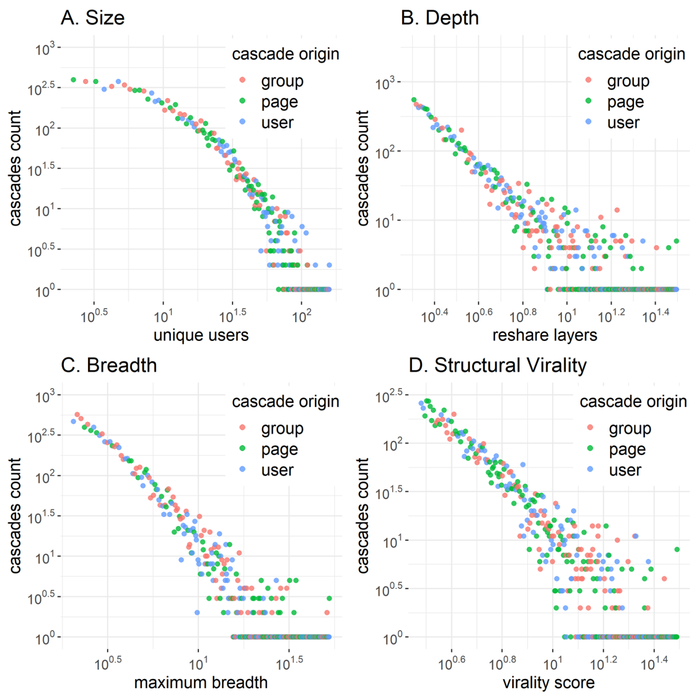

Figure 4: Diffusion Cascades across User Types. Random synthetic data. Panels A-D will summarize the statistical properties of diffusion trees started by groups, pages, and individual users. The aim is to determine if there are significant differences in the reach of content depending on the identity of the root account posting it (no differences in this random data).

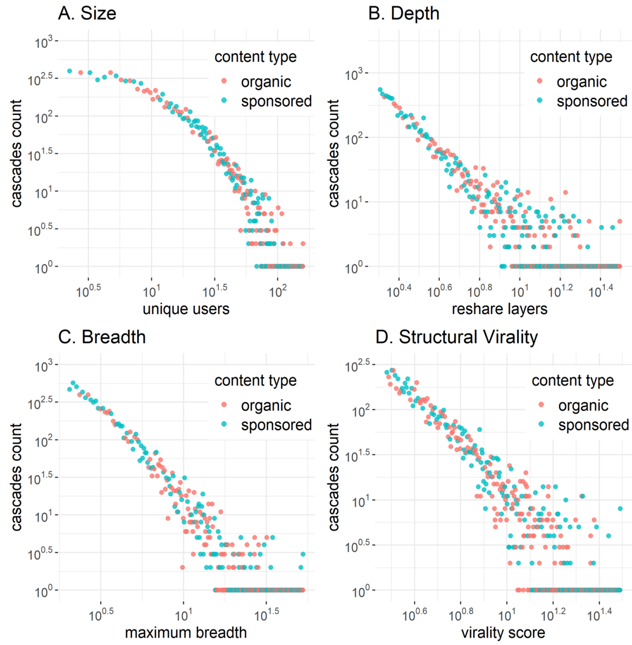

Figure 5: Diffusion Cascades across Type of Content. Random synthetic data. Panels A-D will summarize the statistical properties of trees diffusing organic and sponsored content. The aim is to determine if there are significant differences in the reach of paid ads, compared to non-sponsored posts (no differences in this random data).

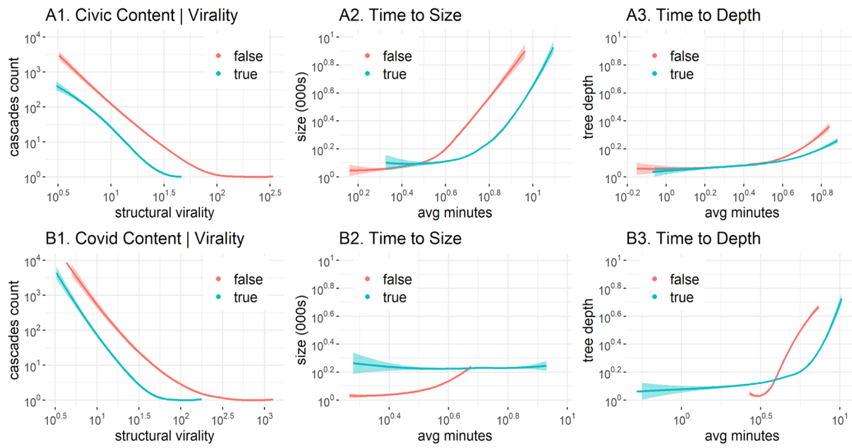

Figure 6: Differences in the Diffusion of True and False Content. Random synthetic data. Panels A1-A3 will summarize the differences in structural virality, time to size, and time to depth between civic content labeled as 'misinformation' and reliable content. Panels B1-B3 will summarize the same differences for Covid-related content. The aim is to determine if false information spreads further and faster (as prior research has documented for Twitter data).

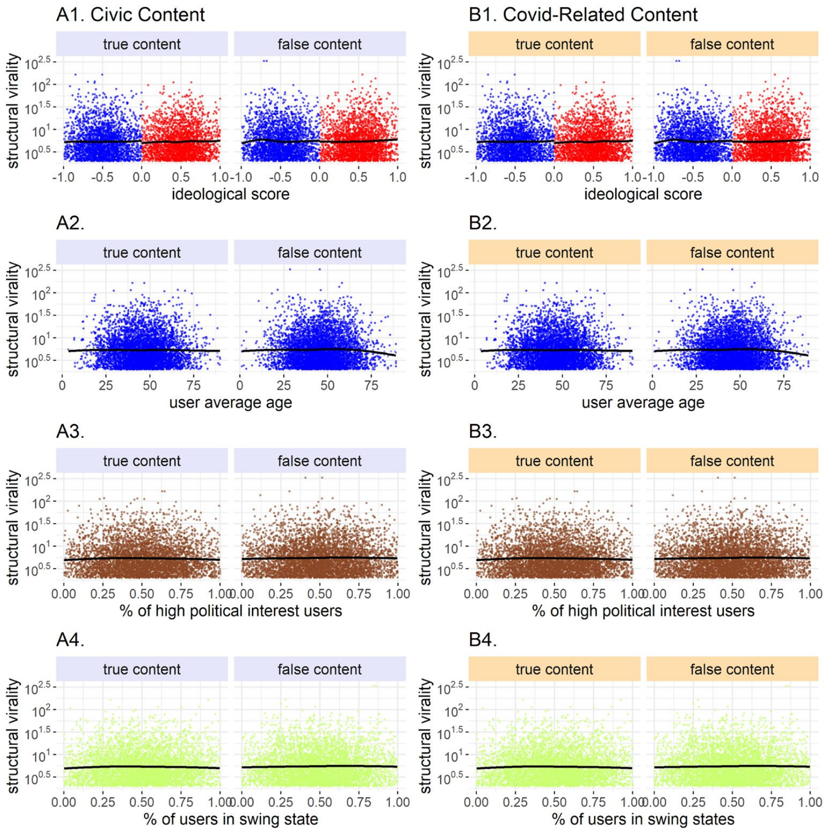

Figure 7: User Composition of Trees Classified as Spreading True and False Content. Random synthetic data. Each data point represents a tree. Panels A1-A4 will look at civic content; panels B1-B4 will look at Covid-related content. 'Ideological Score' measures the fraction of liberal and conservative users in each tree. The plot will allow us to determine if conservatives (or liberals) are more prevalent in the spread misinformation. 'Average age' measures the average age of users in each tree. The plot will allow us to determine if older users are more prevalent in the spread of misinformation. '% of high political interest' measures the fraction of users in a tree that are classified as having high interest in politics. The plot will allow us to determine if political interest correlates with the spread of accurate information. Since the dataset analyzed will have millions of trees, these graphs will be based on a random sample of trees. Alternatively, we will also consider displaying the data as heatmaps.

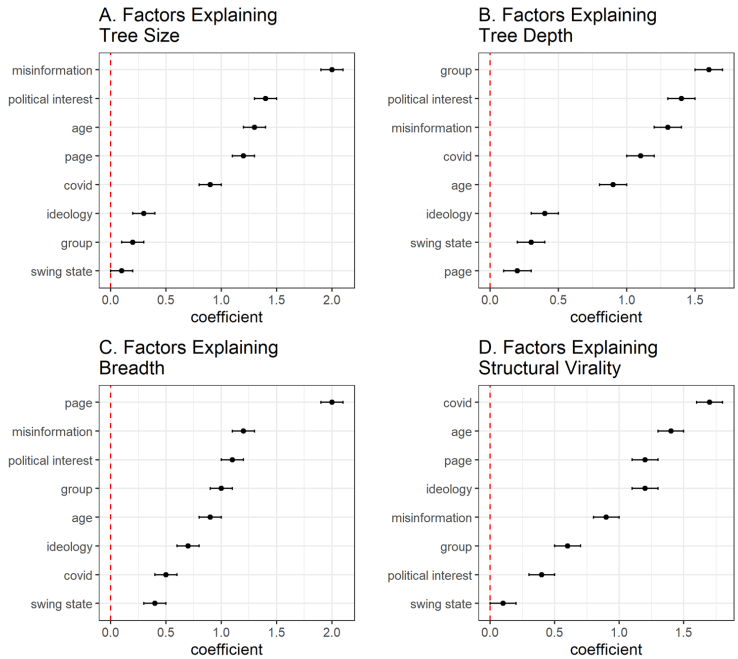

Figure 8: Regression Coefficients Predicting Diffusion Features. Random synthetic data. We will analyze two types of models: four OLS linear regression models with tree size, depth, breadth, and structural virality as dependent variables (the type of outputs for these models are depicted in the figure); and four quantile regression models, with the same dependent variables (these models will be more adequate for the tree data, which will be very skewed, and they will allow us to more directly compare trees in the highest and lowest quantiles. The figure for these models will include estimates for each quantile). Ultimately, the goal of these models is to determine if misinformation is more likely to spread once we control for other confounding factors not considered in prior research (e.g., user age, ideology, level of political interest, and state location).

6. References

1. Liben-Nowell, D. and J.M. Kleinberg, *Tracing Information Flow on a Global Scale using Internet Chain-Letter Data.* PNAS, 2008. 105(12): p. 4633-38.

2. Goel, S., et al., *The Structural Virality of Online Diffusion.* Management Science, 2016. 62(1): p. 180-196.

3. Vosoughi, S., D. Roy, and S. Aral, *The spread of true and false news online.* Science, 2018. 359(6380): p. 1146-1151.

4. Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In *Eighth International AAAI Conference on Weblogs and Social Media*.

S8 Deviations and Clarifications

We disclose the following deviations and clarifications from the registered PAP, along with the justifications for these alterations.

- Clarification #1. Our pre-analysis plan does not include any information about platform interventions or analyses designed to compare tree growth across intervention periods because we did not know at the time of pre-registration that these interventions (and, in particular, the "break the glass" measures described in Figure S11) would be launched and rolled-back at different times and with different degrees of intensity.

- Clarification #2. We added the 'untrustworthy' tree categorization for additional analyses to complement our results on misinformation (see S3.4). These analyses are only reported in the SM.

- Clarification #3. Our pre-analysis plan was created in 2020 prior to the publication of (*6*), which advocates for the need to fix cascade sizes in any comparison of structural properties of diffusion trees. In our analysis, we implemented the matching method proposed by these authors. (See also S3.1).

- Clarification #4. Pages 7-8: we implemented our plan of describing differences across trees categories "for all content and within (...) categories" as a comparison of the pre-registered subset (e.g. misinformation trees) with an appropriate benchmark set (e.g. political trees) and with all trees.

- Clarification #5. Our pre-analysis plan did not consider the possibility of missing branches due to posts created by users outside of the US. Section S3.1 describes the approach we implemented to deal with this issue. This decision was made prior to executing the analyses.

- Clarification #6. Our pre-analysis plan included a graph describing the distribution of post longevity and post shelf-life. In practice, these two metrics were nearly identical, since the only difference was whether the time period starts with the root post or the first re-share, which for large trees tend to be very close in time.

- Clarification #7. Our pre-analysis plan had proposed specific graphical visualizations of the data (see pages 11 and 12). In this manuscript we decided to implement different visualization strategies that better illustrated the patterns we discovered. These changes do not affect the hypotheses we tested or the statistical tests that we had pre-registered.

References

1. S. Goel, A. Anderson, J. Hofman, D. J. Watts, *Management Science* **62**, 180 (2016).
2. S. Vosoughi, D. Roy, S. Aral, *Science* **359**, 1146 (2018).
3. J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, J. Leskovec, Can cascades be predicted? (2014).
4. P. A. Dow, L. Adamic, A. Friggeri, *Proceedings of the International AAAI Conference on Web and Social Media* **7**, 145 (2013).
5. S. González-Bailón, *et al.*, *Science* **381**, 392 (2023).
6. J. L. Juul, J. Ugander, *Proceedings of the National Academy of Sciences* **118**, e2100786118 (2021).
7. A. Friggeri, L. Adamic, D. Eckles, J. Cheng, *International AAAI Conference on Web and Social Media* (2014).
8. R. Chetty, *et al.*, *Nature* **608**, 108 (2022).